Index optimisation for structural equation models (SEM)

Carmen Stindt

Submitted in fulfilment of the requirements for the degree of

Master of Commerce

In the Faculty of Science at the Nelson Mandela University

April 2018

Supervisor: Prof G.D. Sharp

Co-Supervisor: Prof M. Mey

Declaration

Full Names: Carmen Stindt

Student Number: 212256483

Qualification: Magister Commercii

Title of Project: Index optimisation for structural equation models (SEM)

Declaration:

In accordance with Rule G5.6.3, I hereby declare that the above-mentioned treatise / dissertation / thesis is my own work and that it has not previously been submitted for assessment to another University or for another qualification.

Carmen Stindt

Dated: March 2018

Acknowledgements

My sincere gratitude to:

- My supervisor, **Prof G.D. Sharp**, for his guidance support and encouragement throughout this project.
- My co-supervisor, **Prof M. Mey** for her assistance and support.
- **Ross Michael Dix-Peek** for his love, support and keeping me sane during this time.
- My colleagues and friends at the university for their input and support.
- My close friends and family for their support.
- **Ms Rae Vincent-Le Roux** for her administrative help and always being willing to help.
- **The South African Statistical Association (SASA)** for their financial support.
- **Nelson Mandela University** for their financial support.

The financial assistance of the **National Research Foundation (NRF)** towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Abstract

Structural equation modelling (SEM), a statistical technique used extensively in quantitative marketing research and other domains, is an analytical approach used to model latent (unobservable) variables. Unlike distribution fitting where simple chi-squared goodness-of-fit assessment yields satisfactory results, model fit in SEM is more difficult. Descriptive goodness-of-fit indices have been developed over the past 50 years to assist in the assessment of model fit. The traditional assessment method requires reporting multiple indices, all of which should reflect an adequate model fit in order for the overall model fit to be deemed good. The choice of indices to report are left to the researcher's discretion, leading to the indices used to differ considerably. The combination of using the traditional assessment method and differing indices often lead to conflicting results.

This study proposes a composite index, combining frequently used indicators in an attempt to obtain a single index method for assessing model fit in SEM that performs better when compared to the traditional assessment method. Composite indices have been used in other domains as an improved method of assessing performance (Barr and Kantor, 2004).

The composite index proposed is evaluated using a Monte Carlo simulation study under different experimental conditions. The experimental conditions investigated are sample size, estimation method and model misspecification. These experimental conditions are chosen to investigate as each has been shown to affect the traditional indices performances. The ideal fit indices should be able to detect model misspecification while being insensitive to sample size and estimation methods. This is not always the case with the traditional indices.

The composite index proposed is shown to outperform the traditional assessment method under many of the experimental condition combinations. This provides evidence that composite indices may be a more beneficial method of assessing model fit in SEM.

Keywords: *Goodness-of-fit indices, composite index, model fit, structural equation modelling, SEM*.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background of structural equation models (SEM)

Structural equation modelling (SEM) is a statistical modelling technique that is extensively used in the social and behavioural sciences to analyse latent constructs (Bollen and Noble, 2011). This form of analysis is used to quantify behaviours or concepts, such as motivation, quality of sleep and experience of pain (Bollen and Noble, 2011). This analysis provides an opportunity for quantitative measures to be used in areas where traditionally only qualitative research had been used.

The development of SEM began in the early 1900's. As SEM is a combination of several techniques, it does not have a single point of origin. The development of exploratory factor analysis (EFA) by Charles Spearman in 1904 formed one part of the development of SEM (Kline, 2016). In 1918 geneticist Sewell Wright invented the basis for path analysis to aid his work in genetics. He was able to demonstrate how the direct and indirect effect parameters relate to estimated covariances. From this he developed the methodology in which the direct and indirect effects could be estimated from sample data (Kline, 2016). The original SEM models weren't popular until they were introduced into the sociological domain by authors H. M. Blalock in 1961 and O. D. Duncan in 1966 (Bollen and Noble, 2011). The SEM models then grew in popularity in the other social sciences and psychology. The techniques EFA and path analysis, including regression analysis, were first integrated in the early 1970's by K. G. Jöreskog, J. W. Keesling, and D. Wiley, into a model named the JWK model, the precursor to the SEM framework (Kline, 2016).

The first available computer software developed to analyse SEM models was LISREL, based on the LISREL model syntax, developed by K. G. Jöreskog and D. Sörbom in 1973. The name LISREL is an acronym for LInear Structural RELations. Originally, SEM models were restricted to linear models which is no longer a requirement (Hox and Bechger, 2007). This was deemed as the "turning point" in the development of SEM (Bollen and Noble, 2011). In the 1980s and 1990s the development of specialised software occurred along with the rapid expansion of the use of SEM techniques within the social sciences (Kline, 2016). The computer programs developed include AMOS, Mplus and EQS. From this point, SEM analyses started to move into other domains such as biostatistics and epidemiology (Bollen and Noble, 2011). With the development of SEM software, these models have become easier to use.

With the increased interest in Bayesian methods for analysis in the social sciences, Bayesian methods for SEM were developed and integrated into software capabilities such as AMOS and Mplus (Kline, 2016). This further expanded the capabilities of SEM analyses and the types of applications. Further developments included the addition of bootstrapping abilities within the various software. In closing

SEM applications grew in domain, from social sciences to epidemiology, as software development provided modelling opportunities.

1.2 Problem statement and objectives

An integral part of any statistical analysis is the opportunity to assess how well the data fits the selected model. Arguably the most well-known methods are the coefficient of determination R^2 used in linear regression and the χ^2 goodness-of-fit test used in distributional assessment (Iacobucci, 2009). In SEM, the ability to assess model fit becomes challenging due to the nature of the data traditionally used. The traditional method used to assess model fit in SEM was the *p*-value associated with the likelihood ratio χ^2 statistic (Gerbing and Anderson, 1992). This statistic tests, under the null hypothesis, that the theoretical covariances are not significantly different from the estimated covariances from the correctly specified model. A correctly specified model should therefore lead to an insignificant *p*-value. However, the χ^2 statistic, and thus the *p* value, is affected by sample size. This led to situations where good-fitting models were rejected based on the *p*-value. This issue sparked the development of several descriptive goodness-of-fit indices for use in SEM. Well-known indices include, the goodness-of-fit index (GFI), the χ^2 degrees of freedom ratio (CMIN/df) and the root mean square error of approximation (RMSEA).

The problem associated with the goodness-of-fit indices is that there is no universal agreement on which index is the best and should be used (Gerbing and Anderson, 1992). Therefore, most researchers suggest reporting several goodness-of-fit indices to assess model fit. However, there is a lack of agreement amongst researchers regarding the goodness-of-fit indices to be used for model choice (Coughlan, Hooper & Mullen, 2008). For example, according to Hair, Black, Babin and Anderson (2014) three to four indices should be used to evaluate model fit. They suggest that researchers should report at least one incremental index (CFI, TLI, NFI or RNI), one absolute index (GFI, RMSEA or SRMR), and the CMIN/df. In Byrne (2010) the choice of index is more specific in that the CFI, the RMSEA and the CMIN/df ratio is emphasised. Other researchers provide different perspectives. Iacobucci (2009) states the SRMR index should be used in place of the RMSEA index while Weston and Gore (2006) indicate that the CFI, RMSEA, SRMR, GFI and the χ^2 statistic should be reported.

The lack of agreement regarding which indices to report could allow for the situation where researchers only report indices that support their model, ignoring the indices that show an inappropriate model fit. This leads to inconsistent uses of SEM analysis and can lead to contradictory results. The need to report multiple indices also provides an added difficulty, as for the model to be deemed a good-fitting model, all the indices reported should represent this result. This means that should one index reported fail to adequately assess a model, a good-fitting model can be deemed as a bad-fitting model when it shouldn't be classified as such.

This lack of uniformity amongst researchers is the catalyst for this research study. This study aims to investigate whether the use of a composite goodness-of-fit index will perform better in classifying the model fit, when compared to the traditional method of using multiple indices. This is done by developing a composite index from several traditional indices. The choice of indices is based on two components. These components are the frequency of use of each index in the literature and the properties of each index. The composite index is developed using four indices, two absolute indices and two incremental indices. The development of the composite goodness-of-fit index is done by completing the following objectives:

- Reviewing the properties of each index and identifying the critical indices to be included in the composite index. This will be done through the literature review.
- Identifying the most frequently used indices in the applied SEM literature.
- Proposing and defining the composite index.
- Performing a simulation study to assess the performance of the composite index compared to the traditional combinations used to assess fit, under varying experimental conditions.
- Apply the composite index to case study data.

The composite index developed in this study aims to make the following contributions to the SEM domain:

- Assess the overall performance of the composite index for different estimation methods, maximum likelihood and generalised least squares, within a simulated environment.
- Assess the overall performance of the composite index for varying sample sizes within a simulated environment.
- Assess the overall performance of the composite index for two misspecified models, within a simulated environment.
- Compare the performance of the composite index and the performance of the traditional index combinations.
- Use the composite index to assess model fit for two case study models that make use of ordered and randomised questionnaires.

The comparison of the composite index and the traditional method is used to identify whether the composite index could be used as an alternative index to assess model fit. Identifying a better performing index for model fit assessment may lead to an improved SEM analysis. It may reduce the level of confusion amongst researchers regarding the method of assessing model fit to be used.

1.3 Outline of dissertation

This study consists of seven chapters. A summary of each chapter is described below.
Chapter 2, Literature Review, introduces and reviews research required to understand the purpose and methodology of this study. It introduces the theory of questionnaire design, structural equation modelling (SEM), typical software and Monte Carlo simulation studies.

Chapter 3, Composite Index Development, introduces the theory relating to the goodness-of-fit indices developed for assessing model fit in SEM. It also reviews the index frequency from the literature and the subsequent development of the composite index.

Chapter 4, The Simulation Methodology, describes the procedures used to obtain the results. The method used to perform the simulation study is discussed in detail.

Chapter 5, Results and Discussion, contains all the simulation results obtained in the study. The performance of the composite index is analysed and discussed.

Chapter 6, Case study, introduces the theory relating to exploratory factor analysis (EFA) and talent management strategies. It then describes the methodology used to conduct the case study and the results of the case study. Finally, the performance of the composite index and traditional indices is discussed and the results of the ordered and randomised questionnaires are compared.

Chapter 7, Conclusion, is the overall conclusion of the study. It summarises the results, discusses the limitations of the study and possible improvements for future work.

Chapter 2 Literature Review and Theoretical Concepts

2.1 Introduction

The following chapter is a brief review of the literature published on structural equation modelling (SEM) as well as the processes used within the study. Section 2.2 investigates the theory of SEM including the history of SEM (section 2.2.1), the SEM process (section 2.2.2), SEM notation (section 2.2.3), the goodness-of-fit indices used to assess model fit (section 2.2.4) and finally the limitations and misconceptions of SEM (section 2.2.5). Section 2.3 reviews selected software options available for SEM analysis. Section 2.4 introduces the theory of Monte Carlo simulation studies and the process to be implemented to run a Monte Carlo simulation study for SEM models. Finally, section 2.5 reviews the literature on questionnaire design, and comparing ordered and randomised questionnaire designs.

2.2 Structural equation modelling (SEM)

Structural equation modelling (SEM) is a unique statistical technique that can model multiple-equation models (Bollen and Noble, 2011). It allows for the measurement of latent variables that are defined by multiple measures. The advantage of being able to assess latent variables is that it provides researchers with a method of analysing complex theoretical constructs that were previously not investigated due to their unobservable nature. It has the added benefit of being able to handle measurement error, a trait that analysis such as multiple regression, fails to do (Bollen and Noble, 2011). Failure to consider measurement error within an analysis can lead to inaccuracy in the results (Bollen and Noble, 2011). The purpose of SEM is to provide a summary of the interrelationships among variables and the hypothesised relationships between latent factors (Weston and Gore, 2006).

SEM models can be distinguished by three characteristics (Hair et al., 2014). Firstly, it involves the estimation of multiple and interrelated dependence relationships. SEM can represent unobservable variables in these relationships and account for measurement error while estimating the model. Thereafter, SEM can define a model that explains the entire set of relationships using a path diagram. These characteristics will be discussed in further detail within this section.

The general SEM model has certain special cases that are familiar forms of analysis. These include multiple regression, ANOVA, factor analysis, recursive models and growth curves amongst others (Bollen and Noble, 2011). Multiple regression can be viewed as a SEM model in which there is one dependent variable with multiple covariates, while an ANOVA can be seen as a SEM model where the covariates are categorical. The main assumptive difference between these analyses and SEM regards the structural component in the SEM analyses. The structural component in SEM analyses is brought in as a causal assumption by the researcher (Bollen and Noble, 2011).

The traditional regression models consist of one dependent variable with multiple covariates, while SEMs usually have multiple equations within the model (Bollen and Noble, 2011). Therefore, it is not practical to distinguish between independent and dependent variables as traditionally done, as an independent variable from one equation may also be a dependent variable in another. To avoid confusion, the variables are defined as exogenous and endogenous. An endogenous variable is one that takes the place of a dependent variable in at least one equation (Bollen and Noble, 2011). An exogenous variable is one that only takes on the role of an independent variable but never appears as a dependent variable. Exogenous variables are the only variables that can be correlated with each other.

The variables are further defined in terms of latent and observable variables. Latent variables are defined as unobservable concepts that cannot be measured directly and are therefore, represented by several observable variables (also known as indicators). Latent variables typically represent theoretical constructs that researchers are interested in quantifying (Hox and Bechger, 2007). These are measured indirectly by the examination of the consistency between multiple measured variables (Hair et al., 2014).

SEM models can be decomposed into two types of models, namely, measurement models and structural models. A measurement model is defined as the individual latent factors that are explained by several observable variables. A structural model is defined as the relationships between the latent variables. An example of a full SEM model is illustrated in Figure 2.1.

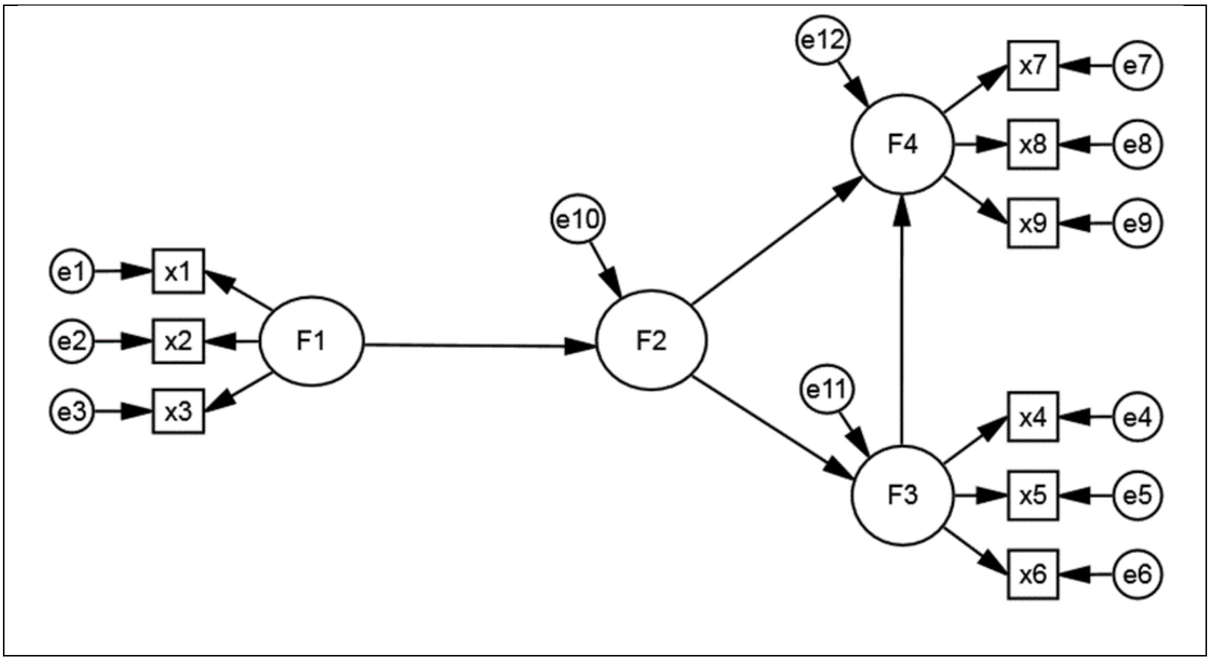


Figure 2.1: An example of a full SEM model

The relationships between the observable variables (denoted x_i, i =1, 2, ... ,9) and factors F1, F3 and F4 represent the measurement models. The relationships between the constructs F1 and F2, F2 and F4, F2 and F3, as well as F3 and F4 represent the structural model (Hair et al., 2014). The measurement error terms are denoted by e_i, i = 1, 2, …, 9, and the disturbance terms are denoted by e_i, i = 10, …, 12.

The SEM technique can be thought of as a combination of factor analysis and path analysis with similarities to analyses such as correlation, multiple regression and ANOVA. There are four main similarities between SEM and these techniques (Weston and Gore, 2006). These include:

- The techniques are general linear models.
- Each technique has underlying assumptions that must be met.
- None of these techniques imply causality.
- Each technique can be misused by researchers if correct procedures are not followed.

An important difference between SEM and these techniques is that SEM allows the researcher to use multiple measures to represent latent factors and model measurement error. The following section will discuss the basics of the SEM process.

2.2.1 The SEM basics

The following sections describe the models and associated assumptions for the SEM models. The general form for the structural model is shown by the following equation (Bollen and Noble, 2011),

$$\eta_i = \alpha_\eta + B\eta_i + \Gamma\xi_i + \zeta_i \quad (2.1)$$

where η_i represents the vector of latent endogenous variables for unit *i*, α_η is the vector of the equations' intercept terms, B is the coefficient matrix for the expected effects of the latent exogenous variables (η), ξ_i represents the vector of other latent exogenous variables, Γ is the coefficient matrix for the expected effects of these latent exogenous variables on the latent endogenous variables, and ζ_i is the vector of disturbances (Bollen and Noble, 2011).

The assumptions for this model include that $E(\zeta_i) = 0$, $COV(\xi_i^{'}, \zeta_i) = 0$, and $(I - B)$ is invertible. The vectors ξ_i and ζ_i should be uncorrelated. Within this general model, the exogenous latent variables are not explained within the model, these variables are explained in the measurement models, while the endogenous latent variables are directly influenced by the variables within the model (Bollen and Noble, 2011).

The measurement models, which represent the relationships between the observed and latent variables, can be represented by the following two equations (Bollen and Noble, 2011),

$$y_i = \alpha_y + \Lambda_y \eta_i + \varepsilon_i \tag{2.2}$$

$$x_i = \alpha_x + \Lambda_x \xi_i + \delta_i \tag{2.3}$$

where y_i and x_i are vectors of observed variables of η_i and ξ_i, respectively, α_y and α_x are vectors of intercept terms, Λ_y and Λ_x represent the matrices of factor loadings for the impact of η_i and ξ_i on y_i and x_i, respectively. Finally, ε_i and δ_i represent the measurement error terms (Narayanan, 2012). The assumptions associated with these equations are that $E(\varepsilon_i) = 0$ and $E(\delta_i) = 0$. The vectors ε_i and δ_i should be uncorrelated with each other.

Using the SEM model as depicted in Figure 2.1, the following diagrams and equations provide an illustration of how the measurement and structural models are mathematically defined. Figure 2.2 depicts the measurement model for latent variable F1 from Figure 2.1. Figure 2.3 shows the structural model for latent variables F2, F3 and F4 from the full SEM model in Figure 2.1.

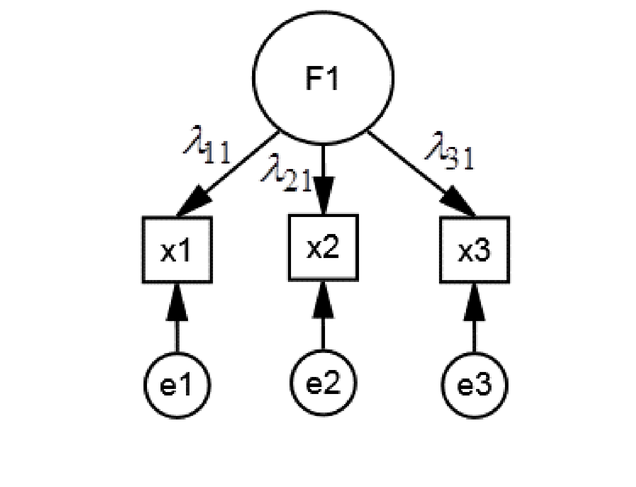


Figure 2.2: Measurement model for exogenous latent variable

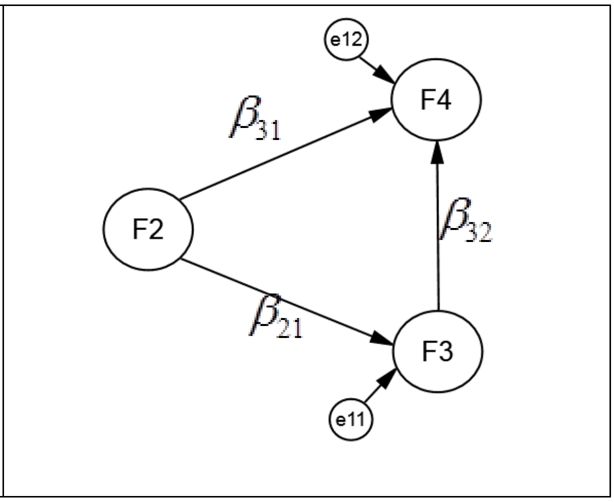


Figure 2.3: Structural model for endogenous latent variables

The model in Figure 2.2 illustrates an exogenous variable (F1) with three observed variables ($x1, x2, x3$) and the corresponding measurement error terms (e_1, e_2, e_3). The relationships between these variables can be defined with the following equations,

$$x1 = \lambda_{11}F1 + e_1 \quad (2.4),$$

$$x2 = \lambda_{21}F1 + e_2 \quad (2.5),$$

$$x3 = \lambda_{31}F1 + e_3 \quad (2.6).$$

The terms $\lambda_{21}$and λ_{31}give the expected differences, also known as the unstandardised factor loadings, in $x2$ and $x3$, respectively, for a one-unit difference in the latent variable, F1. The factor loading associated with $x1$, λ_{11}, is set to one, to allow the remaining pathways to be used as a reference point to scale the latent variable (Bollen and Noble, 2011). This is a common practise in SEM analysis and necessary in order to estimate the factor loadings. Another method of scaling latent variables is to set the latent (exogenous) variables' variance to one.

Figure 2.3 illustrates an example of a structural model based on the full SEM example from Figure 2.1. This model represents the relationships between three endogenous latent variables, based on factors F2, F3, and F4 from the SEM model. Each of these variables can be defined in equation form. Two examples of these latent variables can be defined as follows,

$$F3 = \beta_{21}F2 + e_{11} \quad (2.7),$$

$$F4 = \beta_{31}F2 + \beta_{32}F3 + e_{12} \quad (2.8).$$

The terms β_{21}, β_{31} and β_{32} represent the expected effects, often called the parameter estimates, of the endogenous latent variables on one another. The terms e_{11} and e_{12} represent the disturbance terms associated with the respective latent endogenous variables.

The illustrations in Figures 2.1 – 2.3 provide examples of how measurements and structural models are defined mathematically (Bollen and Noble, 2011). The advantage of using SEM analysis is that all the relationships and variables can be estimated simultaneously. This allows for the complete picture of the conceptual model to be captured within the results. The following section describes the basic steps required to run a SEM analysis.

2.2.2 Basic steps of SEM

The six steps illustrated in Figure 2.4 should be completed when performing a SEM analysis (Kline, 2016). The steps can follow an iterative process, as respecification of models may be required and the process repeated (Kline, 2016). The steps assume that the measurement models (each individual factors) have been checked either using a CFA or are reliable based on theoretical validation.

Step 1: Model specification

Specifying the model is the most critical step in the SEM process (Kline, 2016). It encompasses the process by which the conceptual model is formulated. This provides a visual representation, often in the form of a path diagram, indicating the primary latent variables of interest and the relationships expected between them (Bollen and Noble, 2011). This step also includes the specification of the measurement model, indicating the relationships between observed and latent variables (Kline, 2016). Any correlation between exogenous latent variables should also be defined at this point (Bollen and Noble, 2011). To adequately specify the model, the researcher needs to possess sufficient knowledge concerning the theory and prior research in the area (Byrne, 2010). In certain cases, more than one model is plausible, in which case all models should be investigated. This is important, as once the results are obtained, there is an assumption that the researcher's hypothesised model is correct. Therefore, great care must be taken to specify the model as accurately as possible (Kline, 2016).

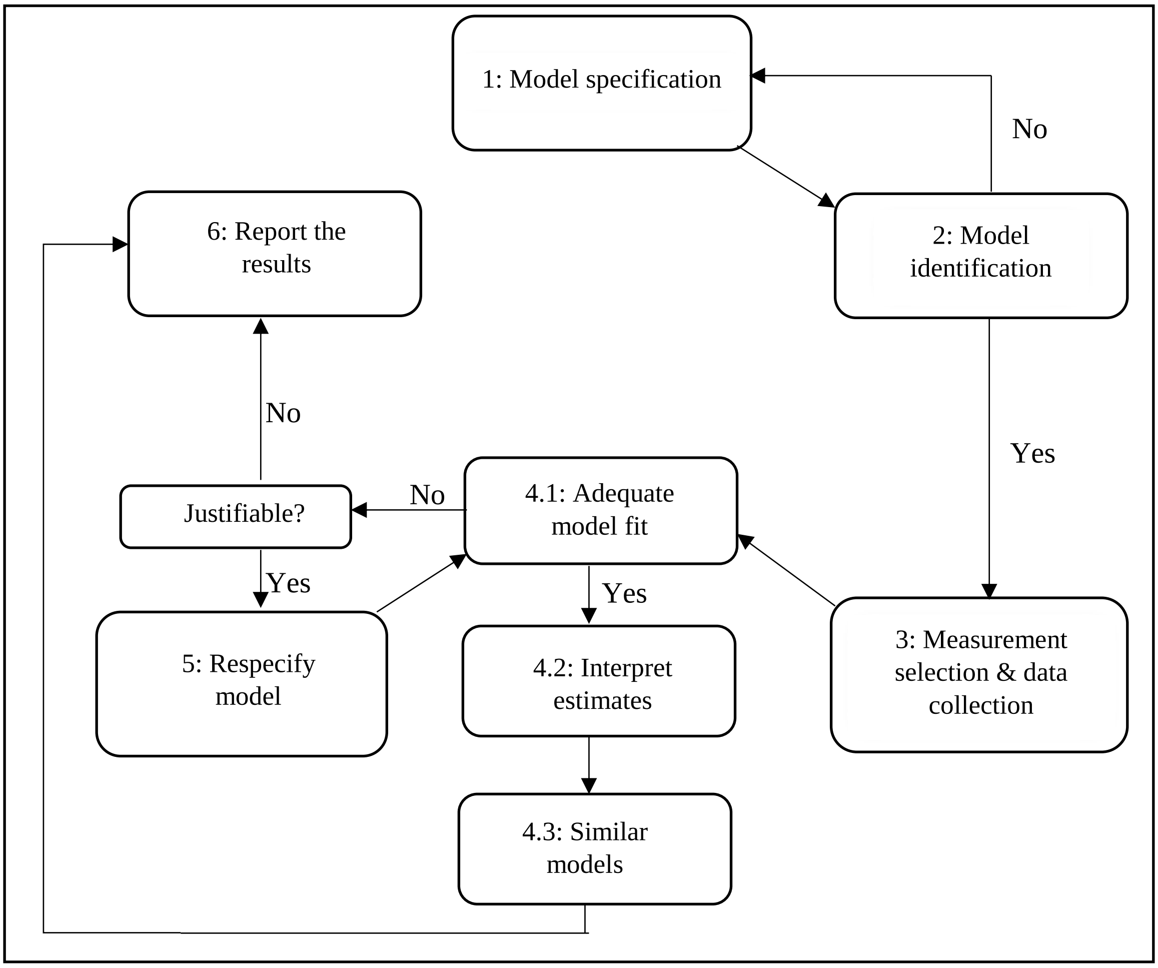


Figure 2.4: General steps within a SEM model

Once the model has been specified, it has an impact on the predictions of moments of the observed variables using the parameters obtained from the model specification (Bollen and Noble, 2011). This is important, as if a model is true, the population and model-implied moments should be close to one another. This is tested within the SEM analysis as it compares the theoretically developed model to empirical data (Nachtigall , Kroehne, Funke, and Steyer, 2003). The analysis minimises the difference between the theoretical model implied covariances and the estimated covariances that are predicted by the model, shown by the following hypothesis,

$$\Sigma = \Sigma(\theta) \tag{2.4}$$

where Σ is the theoretical covariance matrix, θ is the vector of model parameters, and $\Sigma(\theta)$ is the estimated covariance matrix relative to the model parameters. These moments are also important in the estimation and assessment of fit as well as the identification of the model (Bollen and Noble, 2011).

Step 2: Model identification

The SEM process, and relevant software, uses a series of equations and matrix calculations to run the model and obtain the parameter estimates. Therefore, it is important for the proposed models to be identified. A model is said to be identified if it is possible to estimate a unique value for each parameter of the theoretical model. If estimates are not unique, the model is unidentified and cannot be evaluated using the data set (Kline, 2016). A common misconception is that the sample size of the data affects identification. If a model is not identified, obtaining a larger sample size will not influence the identification of the model. Unidentified models require respecification. However, special caution must be taken with regards to the respecification as there must be a theoretical backing to the respecified model. Models should not be respecified until identified without considering whether the new model is logical (Kline, 2016).

Step 3: Measurement selection and data collection

A common approach to collect data for a SEM analysis is via a survey. A researcher will construct a questionnaire based on theoretical underpinnings. Administrating the questionnaire must be done professionally to ensure the integrity of the data. A brief overview of questionnaire design is provided in section 2.5; however, it is important to emphasise that the theoretical development of the questionnaire is crucial if useful data is to be obtained for analysis (Kline, 2016). Once the data has been collected, the cleaning process should be emphasised. The data needs to be screened for any errors to ensure accurate results.

Step 4: Model estimation

Once the data is obtained and the model has been specified, the model is estimated (Kline, 2016). Prior to assessing the model estimation, it is important to consider the four assumptions that should be met in any SEM analysis. These assumptions are:

- Multivariate normality.
- Linearity of all relationships.
- Random sampling of respondents.
- Independent observations.

If these assumptions are met, the researcher can proceed with assessing the model fit and inferential interpretation. Model fit is assessed using a Chi-square test and several goodness-of-fit indices. If the model shows an inadequate fit, a model respecification may be considered. If the respecification is justified, then the new model must be defined and re-estimated. Inadequate model fit during SEM estimation is not uncommon. Hence defining a revision of the original model is common in the literature. This revision must satisfy theoretical requirements in order to be used for further analyses.

Once an adequate model fit is obtained, estimation follows. The SEM analysis estimates the parameters such that the difference between the estimated covariance matrix ($\Sigma(\theta)$) and sample covariance matrix (S) is as small as possible (Hair et al., 2014). This can be done using one of two estimator classes, either model-implied moment (MIM) estimators or model-implied instrumental variable (MIIV) estimators (Bollen and Noble, 2011). The most commonly used estimators are MIM estimators and are based on selecting the model parameters such that they reproduce the moments of the observed variables. These estimators are often called full information estimators as all the relationships within the full model impact on the determination of the estimates (Bollen and Noble, 2011). One disadvantage of using MIM estimators is that if there are any structural misspecifications within the model, the structural error effect associated with the misspecification can affect the full model (Bollen and Noble, 2011). This can influence sections of the model that are correctly specified, leading to potentially misleading results. Some examples of these estimation techniques are maximum likelihood (ML) and generalised least squares (GLS).

The most commonly used estimation technique is ML and is often the default setting is SEM software (Bollen, 1989). The fit function that is used to apply this technique is,

$$F_{ML} = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - p \tag{2.5}$$

where θ is the model parameter vector, $\Sigma(\theta)$ is the estimated covariance matrix as a function of θ, $tr()$ is the trace function, S is the sample covariance matrix, and p is the number of variables. If the assumption of multivariate normality of the observed variables is satisfied, and the model has been correctly specified, the ML technique behaves adequately. The estimator should be asymptotically consistent, efficient, normally distributed and unbiased (Lei and Wu, 2012). In addition, robustness studies done on the ML estimators showed that they maintained their asymptotic properties for continuous distributions that are not necessarily multivariate normal provided they do not have excessive levels of kurtosis (Bollen and Noble, 2011). Arguably this is the reason ML estimation is the most commonly used method. The disadvantage of this method is that it tends to inflate the chi-square (χ^2) test, which can give misleading results by indicating a statistically significant model when a non-significant test is desirable. A non-significant test is desirable as the null hypothesis associated with a SEM analysis indicates a perfect fit. Therefore, a significant test would indicate that the fit is not adequate. Secondly, it decreases the standard error estimates if normality is not observed (Lei and Wu, 2012).

Also available in many software routines is the GLS method of estimation. This method has similar properties to ML as it has desirable large sample asymptotic properties, including unbiasedness (Schumacker and Lomax, 2004). This method also assumes multivariate normality of the observed variables. One aspect to consider is that, as GLS is scale free, transformations made to the observed variables leads to different results between the parameter estimated for the independent transformed and untransformed variables. This method uses the following fit function,

$$F_{GLS} = \left(\frac{1}{2}\right) tr\left(I - \Sigma(\theta)S^{-1}\right)^2 \qquad (2.6).$$

Several other estimation methods, such as unweighted least squares (ULS), asymptotically distribution-free (ADF) and two-stage least squares (2SLS), have been developed to address situations where the models do not meet the assumptions associated with the estimation methods described previously. Further information on these methods can be found in Browne (1984) and Schumacker and Lomax (2004).

The second class of estimators that can be used are MIIV estimators. These estimators are used to deal with structural errors and model misspecifications (Bollen and Noble, 2011). The MIIV estimators make use of scaling factors that are applied to the exogenous and endogenous latent variable vectors. This allows for the removal of latent variables from the general model. The procedure transforms a model containing latent and observable variables into one that only contains observable variables with composite disturbance/error terms (Bollen and Noble, 2011). The equations within this model uses an instrumental variable (IV) approach to deal with the composite disturbance terms. An example of this estimator is the MIIV two-stage least squares (MIIV-2SLS). This method first finds the MIIVs and then applies a 2SLS estimator to the equations. This estimator is more robust than the traditionally used MIM estimators and are helpful to use when structural misspecification is probable (Bollen and Noble, 2011).

Regardless of the estimation method chosen, each method follows an iterative procedure that continues until the difference between the theoretical and estimated covariance matrices is minimised. This ensures that the proposed estimates are as close as possible to the observed estimates in a simultaneous system (Suhr, 2006). This produces the value of the fitting function which is a measure of the degree of correspondence between the observed and expected covariances.

Once the model parameters have been estimated the inferential assessment can proceed. During the inferential assessment the following circumstances should be noted as it is indicative of a misspecification or error in the model (Byrne, 2010),

- Standardised coefficients that are very close to or greater than 1.
- Standard errors greater than |4.0|.
- Insignificant or negative error variances associated with any factor.
- Illogical parameter estimates.

Models with results of this nature will require further investigation and possible model respecification.

The final step in the model estimation process is to consider similar models that would explain the data just as well as the proposed model and the researcher needs to motivate why the model proposed should not be rejected for one of the other models possible (Kline, 2016).

Step 5: Respecification

The process of respecification is a common step as very few SEM models are specified perfectly from the beginning. Therefore, researchers will often need to respecify the models, not only based on statistical properties, but more importantly on the theoretical rationale for the respecification (Kline, 2016). Modification indices (MI), also known as the Lagrangian Multiplier test statistic, is a popular empirical method used to guide model respecification. The MI indicates the change in the chi-square (χ^2) statistic by freeing the pathway of a parameter that was originally fixed (Bollen and Noble, 2011). Only using MI to guide model respecification can lead to illogical estimates therefore, it is important to use MIs in conjunction with theoretical reasoning.

Step 6: Reporting the results

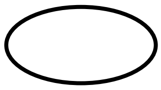
Once an adequate model has been obtained and retained, the results are reported. There are specific guidelines regarding what to report and how to adequately describe the results. These guidelines can be found in work by Boomsma, Hoyle and Panter (2012), MacCallum and Austin (2000), and Thompson (2000). The reason the method of reporting results is important is to allow the researcher to draw the correct interpretations from the model, and therefore provide correct information to readers (Kline, 2016).

An additional two steps that can be employed in the SEM process include replicating the results and applying the results to situations (Kline, 2016). These steps can only be implemented correctly if each of the previous steps were performed appropriately. The following section describes the notation used to illustrate hypothesised models in SEM analyses.

2.2.3 SEM notation

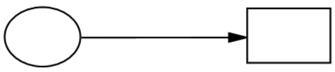
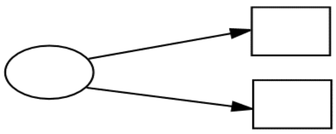
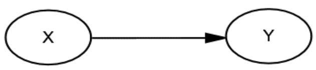
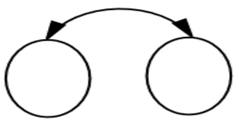
Most SEM models hypothesised are analysed using a path diagram. A path diagram illustrates the relationships between the measured variables and their corresponding latent factors by means of arrows. Table 2.1 shows the graphical elements used in the path diagram along with what each element represents (Hair et al., 2014). SEM analysis can be performed using several softwares including AMOS, LISREL, EQS and MPLUS. These software uses the path diagram prepared by the researcher to run the analysis (Hair et al., 2014).

Table 2.1: Graphical elements used in CFA

Representation	**Element**
	Observed variable
	Latent variable
	One-way relationship
	Covariance relationship

As previously stated, SEM models consist of measurement models and structural models. The measurement model represents the measured variables and how these variables represent the latent constructs (Schumacker and Lomax, 2004). The structural model shows the association of the latent constructs with each other. Both models are represented using path diagrams. Table 2.2 depicts the types of relationships that can be tested in measurement and structural models (Hair et al., 2014).

Table 2.2: Typical types of relationships in a SEM model

Representation	**Relationship Type**
	Relationship between a latent variable and an observed variable (denoted *X* for exogenous and *Y* for endogenous).
	Relationship between a latent variable and multiple observed variables.
	A structural dependent relationship between two latent variables, from *X* to *Y*.
	A correlational relationship between two latent variables.

As SEM models include both measurement and structural relationships, the models developed will likely include both dependence and correlational relationships (Schumacker and Lomax, 2004). The graphical elements and types of relationships illustrated in Table 2.2 are used to describe these models.

2.2.4 Goodness-of-fit indices

An important aspect of any statistical analysis or model, is the assessment of model fit. Model fit in SEM analyses can be divided into two parts (Bollen and Noble, 2011). The first part of model fit regards assessing the component fit of the model. This part is described later in this section. The second part of model fit relates to the overall model fit. This is generally done using tests such as the chi-square (χ^2) test which is available under the full information estimators. This test has a null hypothesis, $H_0 : \mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. Therefore, a good model fit is represented by an insignificant test statistic, indicating that the structure of the estimated covariance matrix equals the population covariance structure (Bollen and Noble, 2011). The chi-squared goodness-of-fit inferential method was found to be inappropriate for SEM in that sample size adversely impacts the test. It was determined that the large sample sizes used in SEM generally ended with analyses implying the model fit was inadequate. This problem was addressed by several researchers, many who proposed alternative model fit methods (Gerbing and Anderson, 1992).

Multiple indices have been developed over the years to assess the fit of SEM models. There is no consensus on a single index that can be used to identify a good model fit. The existing practice is for a researcher to consider several indices to assess the model fit (Schermelleh-Engel, Moosbrugger, and Muller, 2003). Over time these indices have gathered support for a cut-off value that is used as a guideline for assessing goodness-of-fit (Bagozzi and Yi, 1988; and Markland, 2006).

There are three groups of goodness-of-fit indices that can be used. These are absolute fit indices, incremental fit indices and parsimony fit indices (Hair et al., 2014). The commonly used indices can be found in Table 2.3. The goodness-of-fit indices are assessed by comparing its value to a generally accepted cut-off value. Each index has a cut-off value that implies a good model fit, indicating that the model fits the data. These cut-off values have been obtained through extensive simulation studies by the index developers (Hu and Bentler, 1999). It must be noted that the guideline values should be viewed as a general rule of thumb rather than an absolute cut-off point. Cut-off values for adequate fitting models and bad fitting models are also provided in the literature. An adequate fit is generally described as a model fit that is very close to a good fit but does not make the cut-off. The cut-off value for a poor model fit provides a measure that indicates that the model does not fit the data. Any value obtained that is lower than this cut-off (or greater than in certain indices) is interpreted as a bad model fit. Each cut-off value differs slightly depending on sample size therefore the cut-off values listed are for a sample size of two categories.

Absolute fit indices are direct measures of how well a model replicates the original data (Hair et al., 2014). Incremental fit indices are indices that assess how well the estimated model fits relative to a base model. The base model is often referred to as the null model which is a model that assumes that all the observed variables are uncorrelated. Parsimony fit indices provide information about which model is the best model out of a group of competing nested models. Parsimony fit indices are mainly used in the comparison of fit of competing nested models (Hair et al., 2014). These fit indices are explained in detail in the following chapter.

Table 2.3: Goodness-of-fit indices[1]

Index	Index Group	Cut-off value (N < 200)	Cut-off value (N > 200)
Chi-square statistic	Absolute fit index	Insignificant *p*-value	Insignificant *p*-value
CMIN/df	Absolute fit index	< 3.00	< 5.00
GFI	Absolute fit index	> 0.95	> 0.90
RMSEA	Absolute fit index	< 0.08	< 0.08
RMR	Absolute fit index	< 0.08	< 0.08
SRMR	Absolute fit index	< 0.09	< 0.09
NFI	Incremental fit index	> 0.95	> 0.90
TLI	Incremental fit index	> 0.97	> 0.95
CFI	Incremental fit index	> 0.95	> 0.90
RNI	Incremental fit index	> 0.95	> 0.90
AGFI	Parsimony fit index	> 0.95	> 0.90
PNFI	Parsimony fit index	The higher model value	The higher model value

[1] Hair et al., 2014; Byrne, 2010; Hu and Bentler, 1999

It is not only important to assess whether the overall fit is acceptable, the estimated parameters for the structural relationships also needs to be assessed (Byrne, 2010). The parameter estimates need to be statistically significant, the standard error needs to be within an acceptable range and the estimates need to be logical. The parameter estimates are deemed logical by reviewing the sign of the estimate. If the theory states that there should be a positive relationship between two constructs and the parameter estimate for that relationship is negative, the estimate is illogical (Hair et al., 2014). The researcher can then investigate why this empirical result is obtained. In addition the Wald test can be used to test whether the parameters (or groups) of variables are significant simultaneously (Bollen and Noble, 2011). This multivariate procedure is used to determine whether additional variables should be included in the model (Bentler, 1990).

It is important to remember that a proposed theoretical model with an acceptable fit may not be the only model with an acceptable fit. Therefore, it is important to investigate the results of the model with the best fit and comparing it to nested or equivalent models (Schermelleh-Engel et al., 2003). It is equally important that the model is theoretically justified. A model is considered nested within another model if it contains the same or a reduced number of variables and is formed by either deleting or adding relationship paths. All the information in the nested model is found within the larger model. This comparison is done by testing the chi-square difference statistic ($\Delta\chi^2$) calculated by equations (2.8) and (2.9)

$$\Delta\chi^2_{\Delta df} = \chi^2_{df(B)} - \chi^2_{df(A)} \quad (2.8),$$

$$\Delta df = df(B) - df(A) \quad (2.9),$$

where B relates to the complete model and A refers to the nested model. This comparison is used to ensure that the most parsimonious model is chosen. The next section discusses the limitations and misconceptions of SEM analysis.

2.2.5 Limitations and misconceptions of SEM

While the development of SEM has been a big influence in developing quantitative methods for the social sciences (Hox and Bechger, 2007), it has also led to a number of misconceptions regarding the implementation of analysis. This has been as a result of the developments of user-friendly SEM software, such as AMOS, LISREL, EQS, R and MPlus. These softwares have made it increasingly easy to apply SEM models to any dataset, as most software allows models to be drawn using path diagrams (Hox and Bechger, 2007). Therefore, the researcher does not need to understand the intricacies of the mathematical calculations to run the model. This lack of understanding can lead to the violations of the underlying assumptions of SEM analysis.

Another limitation in SEM analysis is that models are not confirmed by data (Cliff, 1983), the data rather fails to not confirm the model. This means that while a model may show a good model fit, it does not provide evidence that the model is theoretically correct. A good fit represents that the estimated and sample covariance matrices are not significantly different, a result which can occur with multiple models (Cliff, 1983).

A misconception in the implementation of SEM is the demonstration of causality. Causality can only be demonstrated through the active control of variables (Cliff, 1983). Causality is often implied by researchers when in fact the relationship is correlational. This lack of control of the variables is due to the nature of the data used in the social sciences (Cliff, 1983).

While there are several limitations to using SEM analysis, there are several advantages to using SEM (Nachtigall, et al., 2003). The benefits of using SEM is that it allows for the analysis of latent variables. This is useful in the social sciences as it allows researchers to analyse the dependencies of latent variables while taking measurement error into account. Using a repeatable measurement design also allows for the calculation of the true-scores for the latent variables through repeating the study in different geographical locations and at different moments in time. Finally, factor scores can be estimated using SEM analysis and used in subsequent analyses (Nachtigall et al., 2003).

Some common errors by researchers using SEM are discussed in the following paragraph. Firstly, most researchers do not adequately translate their important theory into a workable model. A SEM model is only as good as its theoretical basis. Secondly, there needs to be a sufficient number of indicators or observed variables per latent variable in order for the model to be identified (Nachtigall et al., 2003). Thirdly, the modification indices (MI) should not be used purely to improve model fit, the MI used should be backed by theoretical evidence. Researchers could be tempted to include all pathways with significant MIs to improve the model fit without reviewing whether the relationships are logical. The final error commonly made by researchers is that they fail to clean the data adequately. Uncleaned data can lead to misleading results (Nachtigall et al., 2003).

2.2.6 Conclusion

In closing SEM is a unique multivariate analysis that provides researchers the ability to model latent and observed variables. Significant research has been completed in this area in recent years, however, there are still aspects of SEM that require further study. One of these aspects regards the best method of assessing model fit. This study aims to add to the field of SEM research by assessing whether a composite index may perform better than the traditional methods. It will do this by proposing a composite index (chapter 3) and comparing its performance to the traditional indices (chapter 5).

2.3 SEM software

Due to the development of SEM software in recent years, it has become increasingly easy to estimate a SEM model. There are several software packages available to researchers and a brief overview is provided to the intended reader.

Software available to researchers includes AMOS, LISREL, EQS, STATA, MPLUS, Statistica and SAS PROC CALIS. Several packages have also been developed in the freeware R, namely, sem, lavaan, and OpenMx (Paxton et al., 2001). These packages allow users to estimate parameters of a model and calculate various goodness-of-fit indices. Each software package has certain capabilities for running SEM models (Narayanan, 2012). These capabilities include the ability to deal with single groups, multiple groups, missing data, and non-normal variables. Most of these capabilities are consistent across the packages however the biggest difference is whether it contains a graphical interface or not.

An important aspect in analysing any SEM model is assessing the model fit. This is done through the assessment of goodness-of-fit indices therefore, when selecting a software for SEM, it is important that

it contains sufficient indices. The software LISREL, AMOS, SAS PROC CALIS and EQS tend to report most of the available indices (Narayanan, 2012). The R packages and MPLUS tend to report fewer indices however, the user is able to calculate the remaining indices from the data provided.
When the software packages are compared relative to the performance analysing single group models, the results are consistent across the packages (Narayanan, 2012). The main difference observed is in the error estimates, which is dependent on the default estimation method of each package. Therefore, the default estimation method needs to be considered for consistent results to be obtained. The handling of multiple-group analyses involves testing for measurement invariance across groups (Narayanan, 2012). All the software mentioned, excluding the R package sem, can handle multiple-group analyses.

A common occurrence in statistical studies is missing data. The methods of the software packages for handling missing data includes listwise and pairwise deletion, full information maximum likelihood (FIML) method, and multiple imputation (MI) methods (Narayanan, 2012). All the software mentioned previously can use the FIML method. The MI method creates multiple data copies by replacing the missing values in each copy with different estimates. These are then analysed separately and the estimates are pooled into a single value. There are different options within the software packages regarding the stages of imputation, analysis, and pooling. MPLUS offers options to complete all the stages of MI, while AMOS and LISREL have options relating to the imputation stage (Narayanan, 2012). The analysis and pooling stages are not provided in AMOS or LISREL. An important consideration when dealing with missing data is that certain goodness-of-fit indices cannot be calculated due to the indices' dependence on the saturated and independence models. This may affect the researcher's method of assessing model fit adequately.

In conclusion, each of the SEM software packages discussed tends to be consistent regarding results with the main difference being the availability of specific options (Narayanan, 2012). Therefore, the choice of package used will depend on the type of model being analysed. Other concerns to consider when selecting the software package, is the researcher's ability to code, and the budget available for software. If the researcher is not comfortable with coding, then using software that has a graphical interface is advised instead of software that uses a syntax method. Finally, the cost of the packages varies from the freeware R to the more expensive software SAS, LISREL, STATA or MPLUS.

During this study, the opportunity to use R, Statistica, and AMOS were available. Therefore, the preferred options for analysing case study data is AMOS as it reports most of the indices, handles single and multiple group analyses and makes use of both syntax and graphical methods to run the analysis. For simulation studies, the preferred option is R due to its ability to change experimental conditions required in these types of studies and the fact that it is freeware. It also provides the common estimation methods and goodness-of-fit indices. AMOS and R software are used in this study.

2.4 Monte Carlo simulation studies

One of the many areas of research in the SEM domain is the assessment of model fit. As discussed in section 2.2, model fit is assessed using several goodness-of-fit indices. The investigation of the behaviour of these indices under different conditions is primarily done through Monte Carlo simulation studies (Gerbing and Anderson, 1992). The first study of this kind was reported by Boomsma in 1982 and 1983, studying the behaviour of the χ^2 statistic under varying conditions. The first study to assess a descriptive goodness-of-fit index was reported in the work by Bearden, Sharma and Teel (1982). In this study, the NFI's behaviour, under varying conditions, is investigated. These studies are followed by more comprehensive studies such as reported in Anderson and Gerbing (1984). Monte Carlo simulation studies and their use in SEM will be discussed in the following sections.

2.4.1 Monte Carlo simulation studies

Monte Carlo simulations studies are traditionally used to investigate distributional properties of random variables. This is done using simulated random datasets to create sampling distributions to evaluate the behaviour of specified statistics (Paxton et al., 2001). While Monte Carlo studies are a useful technique in the absence of asymptotic theory, it does have several limitations (Gerbing and Anderson, 1992). The results of a study generally represent a compromise as few studies investigate all possible models or conditions. This is due to the vast number of possibilities that can be investigated. Another limitation is that any statistic calculated from the sampling distributions generated within the study are sample statistics. Therefore, these statistics approximate the population values (Gerbing and Anderson, 1992). The advantage of using Monte Carlo simulations is that it allows the researcher to control conditions when estimating the sampling distributions. This provides a vast number of investigative possibilities within each study.

2.4.2 Monte Carlo simulation studies in SEM

Monte Carlo simulation is a popular technique used within SEM research. It is primarily used in the investigation of the behaviour of estimators and goodness-of-fit indices (Gerbing and Anderson, 1992). For each index studied, empirical sampling distributions are generated to assess the overall performance of the index. The experimental conditions of the study can be varied to assess the index's behaviour. The primary reason for using Monte Carlo simulation studies to assess the index properties in SEM is due to the lack of mathematically derived distributions for the indices (Gerbing and Anderson, 1992). In some cases simulation is the only method of obtaining information on the sampling distributions of models under specific experimental conditions (Paxton et al., 2001). The limitation of Monte Carlo simulation studies in the current literature assessing fit indices, is the number of models considered. It would be ideal to study a large variety of models, since model choice could influence the conclusions reached in the Monte Carlo simulations.

Due to the nature of this study, Monte Carlo simulation studies are used to assess the indices' behaviour under varying conditions. The following section describes the general simulation steps and how they were implemented within this study.

2.4.3 Simulation steps

A Monte Carlo simulation study follows four general steps. The first step of the simulation study consists of creating a model with known population parameters. Using the model created in the first step, samples of a specified size (N) are repeatedly drawn and the parameter of interest is estimated (Paxton et al., 2001). The parameter estimates from each sample are then collected and used to estimate a sampling distribution for the population parameter. The final step is to calculate the properties from the sampling distribution estimated. These properties include statistics such as the mean and variance of the sampling distribution (Paxton et al., 2001). These steps were modified for the use within this study. These modified steps are illustrated in Figure 2.5.

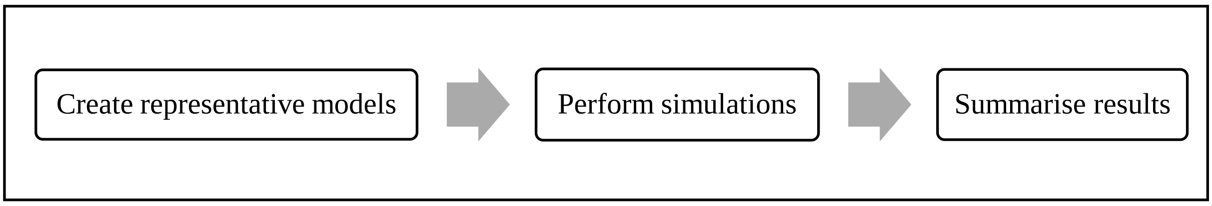


Figure 2.5: Simulation study steps

Step 1: Create representative models

The purpose of this study is to assess the behaviour of SEM fit indices and the composite index, proposed in the following chapter, under varying conditions. The conditions that are varied include the sample size, estimation methods and model misspecification. Therefore, the model used to simulate the data is created taking these conditions into consideration. The model created within the study is based on a previous case study completed by the researcher. The models used in the simulation are in the form of a measurement model as is customary when using latent variables (Hu and Bentler, 1998). The population parameter values selected were obtained from the previous work performed by the researcher.

Step 2: Perform simulations

Once the model has been created, the simulation study is performed. Prior to completing the simulation study, the software used is selected. The simulation study was performed in the software R due to its capabilities discussed in section 2.3. The simulation is done by estimating 200 replications of each population model parameter, for each sample size being investigated. The indices being investigated are then estimated for each sample size. The index estimates are also calculated for the two estimation methods and model misspecifications. During the estimation of each index under these conditions, any samples that failed to converge were removed from the study. Details of this methodology can be found in chapter four.

Step 3: Summarise results

The final step of the simulation study is to calculate the properties of the estimated sampling distribution. The index estimates are captured for each experimental condition and the sampling distribution estimated. The results can be summarised by calculating the mean, variance and frequency of each parameter estimated. This study makes use of means and frequencies. Further details of the simulation study applied in this study can be found in chapter four.

2.4.4 Monte Carlo simulation study conclusion

Monte Carlo simulation studies have become very popular and useful to assess theory by considering properties of estimators and goodness-of-fit indices (Paxton et al., 2001). The general steps used to perform a Monte Carlo stimulation, as well as the study specific steps, have been discussed in the previous section. These steps are used to assess the goodness-of-fit indices being investigated in this study using the simulated data.

2.5 Questionnaire design

A questionnaire is defined as a formalised set of questions used to obtain information about specific concepts from respondents (Malhotra, 2006). Questionnaires are an important scientific instrument as they allow for primary quantitative data collection in a standardised method. This provides internal consistency within the data and allows researchers to obtain coherent analyses (Malhotra, 2006).

The use of questionnaires ensures that the data collected is comparable across respondents. It also increases the speed and accuracy of data collection, allowing for the data to be processed more efficiently (Malhotra, 2006).

2.5.1 Questionnaire design process

A critical aspect that needs to be taken into consideration when using a questionnaire as a data acquisition instrument, is the process of designing the questionnaire or developing a scale (Hinkin, 1998). This often includes the development of a scale(s) that defines conceptual constructs (Hinkin, 1998). Questionnaires can either be designed using previously validated scales or newly-developed scales. The questionnaire design process can be seen as both an art and a science (Malhotra, 2006).

There are a number of steps involved in the development of a scale. The first step involves the creation of items that ensures that the construct being investigated is adequately assessed (Hinkin, 1998). These can be created using either a deductive or inductive methodology. Deductive methodology only uses theoretical foundations to create the items. If done correctly, this ensures content validity but it can be very time-consuming (Hinkin, 1998). Inductive methodology uses categorised descriptions, regarding a certain concept from a sample group, that can be used to develop items for subsequent surveys. This is useful when performing exploratory research but it requires extensive post hoc statistical analysis to refine the items (Hinkin, 1998). Further discussion on the development of items can be found in section 2.5.2.

The next step involves the administration of the questionnaire. To test a newly-developed scale it is important to use data from several independent samples. The size required for each sample will be dependent on the statistical analysis that will be required of the data (Hinkin, 1998). The final step in the development of a scale involves refining the items and testing the reliability of the constructs. This is done through statistical analyses such as exploratory factor analysis (EFA), reliability analysis (Cronbach alphas), and confirmatory factor analysis (CFA). Once developed and tested, the scale can be used in subsequent questionnaires.

Once the scale has either been developed or a previously validated scale obtained, the questionnaire can be refined according to the research question. The following guidelines should be considered when designing a questionnaire:

- The information required from the questionnaire needs to be specified based on a well-defined research question or problem statement.
- The individual questions need to be necessary to the study and, the requirement of multiple questions needs to be considered to ensure that concepts are unambiguous.
- The respondents’ ability to answer the questions adequately needs to be considered.
- The respondents’ willingness to answer the questions should be considered.
- Biographical questions should be kept to only those necessary for the study.

These brief guidelines need to be considered to ensure that the data collected is appropriate for the study and the data analysis.

2.5.2 Question structure

Once the questionnaire design process has been considered, the structure of the individual questions needs to be decided. There are two question structure types that can be used in the development of a questionnaire. These are unstructured questions and structured questions.

- **Unstructured questions**

Unstructured questions are open-ended questions (also known as free-response or free-answer questions) that allow each respondent to answer the question in their own words. This is used to allow respondents to express attitudes or opinions without being confined to predetermined answers, which eliminates predefined bias (Malhotra, 2006). There are several disadvantages in using unstructured questions. These include the increased probability of encountering recording errors, the increased difficulty of assigning data codes, and finally, it adds to the complexity of analysis used.

- **Structured questions**

Structured questions are questions using a set format with a limited number of response options. The questions can either be in the form of a multiple-choice question or a scale question. When using a multiple-choice question, respondents can select more than one response (Malhotra, 2006). This leads to two concerns, namely, selecting the number of alternatives to be included, and the order/position of the options. Each alternative included should be mutually exclusive, and the alternatives should be positioned such that respondents do not select them merely because it occupies a certain position in the list.

A scale is an itemised rating score that can have a number (1, 2, 3, …) or a brief description (strongly disagree, disagree,…) that is associated with each response category. The most commonly used scale is the Likert scale (Rattray and Jones, 2005). A Likert scale is used to measure the degree of agreement with a statement (Malhotra, 2006). Likert scales traditionally are designed as a three-point, five-point or seven-point scale; however, a scale can be created using any number of categories. The Likert scales can either use numbers, descriptions or a combination of the two. The descriptions used often include terms such as strongly disagree, disagree, neutral, agree, and strongly agree.

There are four decisions that need to be made when constructing a Likert scale. Firstly, the number of scale categories that must be chosen. The number of categories should be chosen such that it generates a sufficient amount of variance among respondents (Hinkin, 1998). Secondly, the researcher must decide on whether to use a balanced or unbalanced scale, this refers to the ratio of favourable to unfavourable responses. A balanced scale contains an equal number of favourable and unfavourable responses. An unbalanced scale contains more positive than negative responses, or vice versa. Thirdly, the decision of whether an odd or even number of categories should be used, must be made. Finally, the researcher must decide on whether a forced or non-forced scale will be used. A forced scale is structured such that the respondent must give an opinion as the category, "not applicable or no opinion" is not provided. An unforced scale is one in which the category option "not applicable or no opinion" is provided, allowing respondents to choose whether to give an opinion (Malhotra, 2006). Likert scales are commonly used due to the ease of construction and administration of the scale. It is also relatively easy for the respondents to understand.

Finally, regardless of the type of scale used, special consideration should be taken in the wording of the questions/items. The questions/items should be phrased using simple and unambiguous words. The language used should be familiar to the target respondents (Hinkin, 1998). Items should be consistent in terms of the perspective of the study and should only address a single issue per item (Hinkin, 1998). Leading and biased questions should be avoided to obtain reliable data.

2.5.3 Questionnaire structure

Another aspect to consider is the structure of the questionnaire. This consideration depends on the choice of questions included. The general structure that should be used is to ensure that the initial questions are more generalised, followed by more specific questions. The structure can be further developed by reviewing the type of questions used (Malhotra, 2006).

If the questions are chosen to represent a valid construct, that has been theoretically derived and repeatedly tested, an ordered question structure is a suitable design to use. If, however, the design of the study is empirical and the questions chosen are assumed to be mutually exclusive, a randomised approach is a better choice (Malhotra, 2006). Regardless of the design used, the researcher must review the questionnaire to ensure that the questions are not ordered in such a way as to create bias in the choice of responses.

2.5.4 Theoretical basis

The final aspect to be considered in the design process is to ensure that the study is theoretically based and answers a specified research question(s). This goes together with ensuring that the questionnaire is designed such that it can be repeated multiple times, at multiple occasions (Rattray and Jones, 2005).

2.5.5 Conclusion of questionnaire design

In closing, questionnaires are an important instrument for collecting data in a standardised method. If done correctly it provides the researcher with a consistent method of collecting data. This study makes use of a case study in which the data is collected using a structured and randomised questionnaire. Therefore, the points of consideration discussed in this section were used to direct the construction of the questionnaires. The questionnaires made use of structured questions relating to the research question of the case study, while ensuring that the length of the questionnaire was kept to a minimum.

2.6 Conclusion

This chapter discussed the theoretical elements related to SEM, SEM software, Monte Carlo simulation studies and questionnaire design. The reviewing of questionnaire design, SEM and Monte Carlo simulation studies were reported here as each process was used during the completion of this study. The methodology used for each topic discussed is discussed in the following chapters. The next chapter discusses the development of a composite goodness-of-fit index by investigating the properties of each goodness-of-fit index and their frequency of use in the literature.

Chapter 3 Composite Index Development

This chapter discusses the theory of the composite index proposed in this research. A composite index is an index that is constructed from other indices. This study aims to construct a composite index that performs better when assessing model fit, when it is compared to traditional goodness-of-fit indices. This chapter reviews the theoretical properties of several goodness-of-fit indices used to assess model fit (section 3.1). The frequency of use of the goodness-of-fit indices in the literature is reviewed and discussed (section 3.2) after which the proposed composite index (section 3.3) is motivated.

3.1 Properties of goodness-of-fit indices

The assessment of model fit in SEM is one of the more controversial requirements for practitioners (Curran et al., 2003). Researchers concur that assessing model fit is an integral part of any statistical analysis, however, there is no consensus, in SEM, on how best to do this (Gerbing and Anderson, 1992). Multiple indices have been developed over the past 50 years, attempting to provide a "best index" to use in SEM (Gerbing and Anderson, 1992). As seen in Chapter 2, there are three types of goodness-of-fit indices, absolute fit indices, incremental fit indices, and parsimony fit indices and each type of index tests different criteria of model fit. Several indices have been developed for each of these index types.

Each of the indices developed, were propositioned in an attempt to obtain an ideal index to be used to assess model fit, yet consensus eludes the practitioner. Gerbing and Anderson (1992) state that an ideal fit index should possess three main properties. These properties include the following:

- The index should be restricted to a scale from 0 to 1, where 0 indicates no model fit and 1 indicates a perfect fit.
- The index should be independent of sample size therefore, having a small or large sample size should not affect the index result obtained.
- The index should have known distributional characteristics which allows for easier interpretation and the ability to construct a confidence interval.

No existing index satisfies all these conditions and the conditions themselves are a source of dispute amongst researchers (Gerbing and Anderson, 1992). Fan, Thompson and Wang (2009) states that a goodness-of-fit index should be judged on its relative performance based on the following criteria. Each index should be:

- Insensitive to sample size.
- Insensitive to estimation method.
- Highly sensitive to model misspecification.
- Insensitive to model complexity.

Another criteria used to assess indices' relative performances, is the degree of unbiasedness and variation within the estimate (Fan et al., 2009).

As there is no consensus amongst practitioners as to which is the "best index" to assess model fit, the convention is for multiple indices to be reported. There is however disagreement on which indices to report (Hair et al., 2014). This may lead to researchers choosing indices that reflect a good model fit and ignoring those that state otherwise. The study and development of indices is a prominent area of research in the SEM domain. Several of these studies investigate which indices should be reported to adequately assess model fit. For example, McDonald and Ho (2002) states the commonly reported indices include the NFI, GFI, CFI, and TLI. They also found that researchers tended to report either absolute fit indices or incremental fit indices (McDonald and Ho, 2002). Kline (2016) recommends that the χ^2 value should always be reported along with several descriptive goodness-of-fit indices. The goodness-of-fit indices recommended include the CFI, RMSEA, and SRMR (Kline, 2016). Hair et al. (2014) also recommends reporting the χ^2 value, however, does not indicate which descriptive indices to report, rather advises that two incremental and two absolute indices be reported. This lack of consistency lends validity to the rationale of this study.

In addition to research on the selection of indices, numerous studies have been reported to assess cut-off points indicating good model fit for an index. Hu and Bentler (1999) proposed a Two-Index presentation strategy suggesting that a combination of two indices may perform better. This was suggested as most indices do not perform adequately under all experimental conditions. Using two indices instead of a single index allows for the adequate assessment of more experimental conditions. This study assessed the conventional cut-off criteria versus new alternative criteria to determine performance. In their study, Hu and Bentler (1999), assessed the performance of the traditional indices by reviewing the rejection rates and type 1 and 2 error rates for the conventional cut-off values. Through the preliminary assessment, the SRMR index was found to be the most sensitive to misspecified models, therefore Hu and Bentler (1999) suggested that index combinations should include the SRMR.

Several index combinations' performances were assessed by adjusting the cut-off values to determine the optimal cut-off for each index within the combination. Hu and Bentler (1999) found that the combinations in Table 3.1 produced the best results. While this study indicated that on average a more restrictive cut-off may be necessary for model evaluation, further research is required to test these values under different conditions.

Table 3.1: Two Index presentation strategy rules[2]

Index combination	Combination cut-off criteria
RMSEA and SRMR	RMSEA ≤ 0.06 and SRMR ≤ 0.09
CFI and SRMR	CFI ≥ 0.96 and SRMR ≤ 0.09
TLI and SRMR	TLI ≥ 0.96 and SRMR ≤ 0.09

[2]Hu and Bentler, 1999

Therefore, one can understand the confusion researchers encounter regarding assessing model fit. The choice of index, or indices, used to assess model fit should not be taken lightly. This choice should be made to ensure that the indices used perform adequately under experimental conditions, such as model misspecification and sample size.

The following section discusses the most commonly used goodness-of-fit indices and their respective properties. These are divided into sections based on the type of index, absolute fit indices, incremental fit indices, and parsimony fit indices.

3.1.1 Absolute fit indices

Absolute fit indices are direct measures of how well the model replicates the original data. The indices included in this section are the chi-square (χ^2) statistic, chi-square (χ^2) statistic to degree of freedom ratio (CMIN/df), goodness-of-fit index (GFI), root mean square error of approximation (RMSEA), root mean square residual (RMR), and standardised root mean square residual (SRMR).

- **Chi-square (χ^2) statistic**

The chi-square (χ^2) test statistic is the original method of assessing model fit. This test statistic evaluates whether the estimated covariance matrix and theoretical model implied covariance matrix are equal. As the theoretical covariance matrix cannot be directly calculated, the sample covariance matrix will be used in its place. The formula used to calculate this statistic is

$$\chi^2 = \left(N\left(tr\left(S\Sigma(\theta)^{-1}\right) + \log\left|\Sigma(\theta)\right| - \log\left|S\right| - (p+q)\right)\right) \tag{3.1}$$

where S is the sample covariance matrix (calculated from the data), $\Sigma(\theta)$ is the estimated covariance matrix (calculated based on the model), $tr()$ is the trace function, N is the sample size, p is the number of observed endogenous variables, and q is the number of observed exogenous variables. The chi-square (χ^2) statistic is the only inferential statistic used to assess model fit in SEM (Iacobucci, 2009).

A high value obtained indicates that the theoretical covariance matrix and estimated covariance matrix are significantly different from each other. This indicates that the null hypothesis, stated in Chapter 2, should be rejected, indicating a poor model fit (Iacobucci, 2009). A good model fit is represented by a chi-square (χ^2) statistic close to zero, with the corresponding high *p*-value. A *p*-value greater than 0.05 indicates that the model is compatible with the theoretical covariance matrix as the null hypothesis is not rejected.

There are significant limitations to using this statistic to assess model fit. The first limitation is that this statistic assumes that multivariate normality is present. Another limitation is the behaviour of this statistic to model complexity (Schermelleh-Engel et al., 2003). As parameters are added to a model the χ^2 value decreases which shows an improved model fit. This is problematic as it means that this value can be manipulated to indicate good model fit, leading to misleading results. Finally, this statistic is dependent on sample size. As the sample size increases, the χ^2 value increases, indicating a poor model fit even when the fit is adequate.

- **CMIN/df**

The CMIN/df index is a ratio of the χ^2 statistic to the degrees of freedom of the model. A CMIN/df value of less than 3:1 is associated with a good model fit. For sample sizes of 750 or greater, the ratio of 5:1 is associated with a good model fit (Hair et al., 2014). Highly complex models will influence the ratio value, resulting in a ratio of 5:1 or lower representing a good model fit.

- **GFI**

The goodness-of-fit index (GFI) was originally proposed by Jöreskog and Sörbom in 1984 to produce a stand-alone fit statistic that was less sensitive to sample size compared to the χ^2 statistic (Shevlin and Miles, 1998). The GFI was designed to assess the degree of similarity between the sample covariance matrix (S) and the estimated covariance matrix, $\Sigma(\theta)$. Therefore, it tests the model's performance against a null model, which is a model in which the parameters are fixed to 0 (Schermelleh-Engel et al., 2003). The GFI is calculated as follows,

$$GFI = 1 - \frac{\chi_t^2}{\chi_n^2} \qquad (3.2),$$

where *n* represents the null model (baseline model), and *t* represents the target model. The GFI has a range of 0 to 1 where the higher the value, the better the model fit. A value of 0.90 or larger is generally accepted as indicating an acceptable model fit (Shevlin and Miles, 1998).

Even though the GFI was developed to be less sensitive to sample size, simulation studies have shown that it is has a positive relationship with sample size (Gerbing and Anderson, 1992). A study by Bollen and Stine (1992) found that the GFI was insufficiently sensitive to model misspecification.

- **RMSEA**

The root mean square error of approximation (RMSEA) is an absolute fit index that assesses the estimated discrepancy due to the approximation per degree of freedom (Schermelleh-Engel et al., 2003). The RMSEA is calculated as follows,

$$RMSEA = \sqrt{\frac{(\chi^2 - df)}{df(N-1)}} \quad (3.3).$$

A RMSEA value less than or equal to 0.05 is indicative of a good model fit, whereas an adequate model fit is indicated by a value between 0.05 and 0.08. A RMSEA value between 0.08 and 0.10 indicates a mediocre model fit. Any value larger than 0.10 indicates a poor model fit. A 90% confidence interval can be used to assess the estimated value of RMSEA and is reported in most SEM software (Schermelleh-Engel et al., 2003). A lower boundary that contains 0 indicates an exact fit while a lower boundary less than 0.05 indicates a good fit. This index tends to be relatively independent of sample size and favours parsimonious models (Schermelleh-Engel et al., 2003).

- **RMR**

The root mean square residual (RMR) assesses the discrepancy between the theoretical and estimated covariance matrices after the parameters of the model have been estimated (Schermelleh-Engel et al., 2003). The RMR is often referred to as an overall badness-of-fit measure based on the fitted residuals. The RMR is calculated as follows,

$$RMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \hat{\sigma}_{ij})^2}{k(k+1)/2}} \quad (3.4),$$

where $k = p + q$, s_{ij} represents the elements of sample covariance matrix S, and $\hat{\sigma}_{ij}$ represents the elements of the estimated covariance matrix, $\Sigma(\theta)$. A RMR value that is close to zero indicates a good model fit. The range of the RMR is determined by the scales of each observed variable (Coughlan et al., 2008). Therefore, if the scales of each observed variable vary, interpretation is challenging. This is a limitation to using the RMR as there is no standard cut-off value indicative of an adequate model fit, however on average values less than 0.08 generally indicates a good model fit (Schermelleh-Engel et al., 2003).

- **SRMR**

The standardised root mean square residual (SRMR) was developed to address the limitation to RMR. It does this by standardising the residuals to allow for easier interpretation compared to the fitted residuals. This is because the standardised residuals are independent of the unit of measure of the observed variables (Schermelleh-Engel et al., 2003). The SRMR is calculated as follows,

$$SRMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}\left[\left(s_{ij} - \hat{\sigma}_{ij}\right)/(s_{ii}s_{jj})\right]^2}{k(k+1)/2}} \qquad (3.5).$$

A SRMR value less than or equal to 0.05 indicates a good model fit, while a value of less than or equal to 0.1 indicates an adequate model fit (Schermelleh-Engel et al., 2003). This index is insensitive to sample size and model misspecification but is not available in every SEM software.

3.1.2 Incremental fit indices

The incremental indices assess model fit by reviewing how well the model represents the sample covariances in relation to a null model (Gerbing and Anderson, 1992). The incremental indices that will be discussed include the normed fit index (NFI), Tucker-Lewis index (TLI), and comparative fit index (CFI).

- **NFI**

The normed fit index (NFI) was originally developed by Bentler and Bonett (1980) to assess the ratio of the χ^2 value for the null and target model (Hair et al., 2014). To achieve a NFI of 1, indicating a perfect model fit, the target model shows the best possible improvement relative to the null model (Schermelleh-Engel et al., 2003). A NFI value of 0, indicating a complete lack of fit, suggests that the null model and target model obtain the same minimum fit function. In this case the most parsimonious model is preferred, which is the null model (Schermelleh-Engel et al., 2003). The NFI is calculated as follows,

$$NFI = \frac{\chi_n^2 - \chi_t^2}{\chi_n^2} \qquad (3.6),$$

where *n* represents the null model (baseline model), and *t* represents the target model. The NFI value, representing a good model fit relative to the baseline model, should be 0.95 or greater. Values greater than or equal to 0.90 are indicative of an acceptable model fit (Schermelleh-Engel et al., 2003). A disadvantage of using the NFI is that it is sensitive to sample size therefore it is not consistent for models of all sample size.

- **TLI**

The Tucker-Lewis index (TLI), otherwise known as the nonnormed fit index (NNFI), was developed by Bentler and Bonett (1980) extending the work completed by Tucker and Lewis in 1973. It was developed to account for the sample size problem of the NFI (Schermelleh-Engel et al., 2003). The TLI measures the relative fit of the target model to the null model. The general range of the TLI is from 0 to 1, however, as the index is not normed it can occasionally go slightly over the ends of the range (Schermelleh-Engel et al., 2003). While similar to the NFI, the TLI differs in construct as it is a comparison of the normed χ^2 values for the null and target models. Therefore, the TLI takes, to some extent, the degree of model complexity into account (Hair et al., 2014). The TLI is calculated as follows,

$$TLI = \frac{\frac{\chi_n^2}{df_n} - \frac{\chi_t^2}{df_t}}{\frac{\chi_n^2}{df_n} - 1} \quad (3.7),$$

where *n* represents the null model (baseline model), and *t* represents the target model. As the TLI takes model complexity into account, more complex models are penalised, obtaining lower values, while parsimonious models obtain relatively higher values (Schermelleh-Engel et al., 2003). A good model fit is indicated by values of 0.97 or higher, and an acceptable model fit is indicated by values of 0.95 or greater (Schermelleh-Engel et al., 2003). As the TLI is non-normed, the index value can fall outside the range of 0 to 1. However, if greater than 1, researchers suggest setting the index to one and interpreting it similarly to the NFI (Kenny, 2015). An advantage to using the TLI is that it is one of the indices less sensitive to sample size which is desirable for consistency across models (Schermelleh-Engel et al., 2003).

- **CFI**

The comparative fit index (CFI) was developed by McDonald and Marsh (1990) to ensure that models of small sample sizes are not underestimated, as often encountered using the NFI (Schermelleh-Engel et al., 2003). The CFI is calculated as follows,

$$CFI = 1 - \left(\frac{\max[(\chi_t^2 - df_t),0]}{\max[(\chi_n^2 - df_n),0)]} \right) \quad (3.8),$$

where n represents the null model (baseline model), and t represents the target model (Iacobucci, 2009). The range of the CFI is from 0, indicating no model fit, and 1, indicating a perfect fit. A good model fit is represented by CFI values equal to and greater than 0.95, while values equal or greater than 0.90 indicate an acceptable model fit (Hair et al., 2014). The CFI is one of the most widely used indices as it has many desirable properties. The CFI is relatively insensitive to model complexity and sample size (Schermelleh-Engel et al., 2003).

3.1.3 Parsimony fit indices

Parsimony fit indices are used to compare models to determine the best model from a group. The parsimony fit indices that will be discussed include the adjusted goodness-of-fit index (AGFI), parsimony normed fit index (PNFI), and parsimony goodness-of-fit (PGFI).

- **AGFI**

The adjusted goodness-of-fit index (AGFI) was developed by Jöreskog and Sörbom (1989) to account for the bias that results from model complexity. This is done by adjusting for the ratio of degrees of freedom to the number of observed variables in the model. A less complex model with fewer observed variables and therefore, parameters, will obtain relatively higher values of AGFI (Schermelleh-Engel et al., 2003). As the AGFI deals with the comparison of the χ^2 values for the target and null model, it can be categorised as an absolute fit index. However, many researchers categorise it as a parsimony fit index as it was formulated to account for model complexity to favour more parsimonious models. The AGFI values generally range between 0 and 1, however negative values are possible when models with small degrees of freedom are run with large sample sizes (Schermelleh-Engel et al., 2003). The AGFI is calculated as follows,

$$AGFI = 1 - \frac{\chi_t^2 / df_t}{\chi_n^2 / df_n} \qquad (3.9),$$

where n represents the null model (baseline model), and t represents the target model. The AGFI will approach the GFI when the number of degrees of freedom for the model approaches those of the null model (Schermelleh-Engel et al., 2003). An AGFI value of 0.90 indicates a good model fit while values of 0.85 can be indicative of an adequate fit. While the AGFI was developed to account for the degree of model complexity, it still shows sensitivity to model complexity in simulation studies (Schermelleh-Engel et al., 2003). This sensitivity becomes more apparent at smaller sample sizes. The AGFI has been shown to be sensitive to sample size in several simulation studies (Gerbing and Anderson, 1992).

- **PNFI**

The parsimony normed fit index (PNFI) was developed by James, Mulaik and Brett (1982) as a modification to the NFI, by adjusting it by a ratio of degrees of freedom of the model and null model (Hair et al., 2014). This leads to a downward adjustment of the NFI for complex models. The PNFI can be calculated as follows,

$$PNFI = \frac{df_t}{df_n} NFI \qquad (3.10),$$

where *n* represents the null model (baseline model), and *t* represents the target model. The PNFI has characteristics of both incremental and absolute fit indices but it favours more parsimonious models (Hair et al., 2014). The PNFI should be used to compare two or more models, where the higher PNFI value indicates the better model fit, in this case the more parsimonious fit when compared to other models (Schermelleh-Engel et al., 2003). The PNFI values ranges from 0 to 1.

- **PGFI**

The parsimony goodness-of-fit index (PGFI) was developed by Mulaik et al. (1989) as an adjustment to the GFI. This adjustment was in a form of a parsimonious ratio of the degrees of freedom for the target and null model (Schermelleh-Engel et al., 2003). The PGFI is calculated as follows,

$$PGFI = \frac{df_t}{df_n} GFI \qquad (3.11),$$

where *n* represents the null model (baseline model), and *t* represents the target model. As previously discussed the AGFI was also developed to adjust for model complexity. Both the AGFI and PGFI have a downward adjustment effect on the GFI when analysing complex models, however the PGFI effect is stronger for models with fewer degrees of freedom compared to the AGFI (Schermelleh-Engel et al., 2003).

The most commonly reported goodness-of-fit indices are discussed above, however there are many more indices available to assess model fit. From the discussion, none of the indices investigated possessed all the properties desired from a fit index. Therefore, multiple indices are generally reported, however, with little consensus on which indices to report, the possibility of researchers selectively reporting well-fitting indices, favouring their models, is high.

3.2 Frequency of goodness-of-fit indices

In order to obtain more understanding regarding how model fit is assessed within SEM analyses, the researcher reviewed 182 studies that made use of SEM analysis. The studies reviewed spanned many disciplines in which SEM is commonly utilised. These disciplines include business management, psychology, industrial psychology, economics and construction. Some of the journals reviewed were The South African Business Review, SA Journal of Industrial Psychology, Structural Equation Modeling, South African Journal of Economic and Management Sciences (SAJEMS), and Automation in Construction. Each study is reviewed and the number of times each index reported captured (Figure 3.1). The assessment of the frequency concludes at 182 studies as the researcher felt that the frequencies recorded were representative of the commonly used indices. The reported frequencies provide important information to the reader regarding which indices are commonly understood and available in the majority of SEM software (Shah and Goldstein, 2006). This is used to identify common indices to include in the composite index.

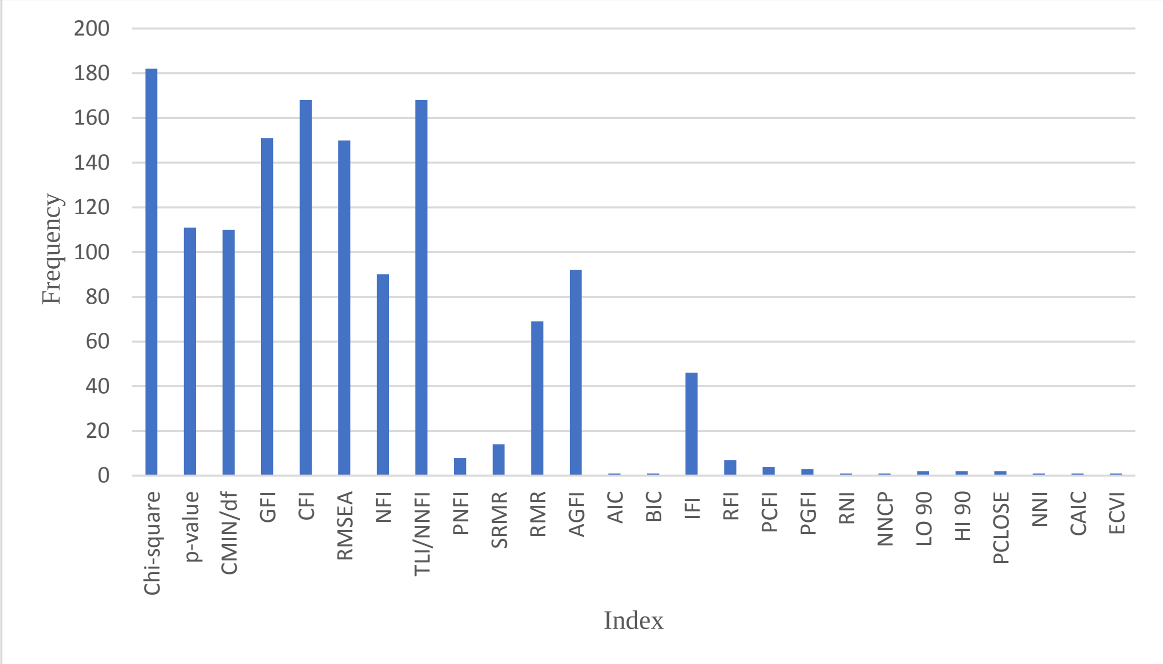


Figure 3.1: Index Frequencies

As seen in Figure 3.1, the most frequently reported method of assessing model fit is the Chi-square (χ^2) statistic. This is consistent with suggestions from Kline (2016) and Hair et al. (2014) regarding the reporting of model fit indices. The second most frequently reported index is tied between the CFI and TLI, while the fourth most frequently reported index was the GFI. The fifth position is held by the RMSEA, the sixth position is held by the p-value associated with the (χ^2) statistic, and the CMIN/df, holds the next position. The parsimony indices PGFI, PCFI, and PNFI were rarely compared, this may be because these indices are commonly used to compare two or more models. The RMR index is reported 69 times while the SRMR is reported 14 times, despite the improved properties of the SRMR. This may be because the SRMR is not readily accessible in all SEM softwares. From Figure 3.1, the spread of absolute and incremental indices is relatively even, which is consistent with the recommendations by Hair et al. (2014) and Kline (2016).

3.3 Composite Index

Composite indices to assess model fit is not a new concept. Gonzalo and Pitarakis (1998) use a linear combination of indices to assess model fit of restricted vector autoregressive models. Gonzalo and Pitarakis' (1998) study assessed the performance of model selection approaches and commonly used penalty types such as Akaike information criterion (AIC) and Bayesian information criterion (BIC). A composite index, linear combination information criterion (LCIC), was introduced, which is a linear combination of two traditional criterion, BIC and Hannan-Quinn (HQ). The reason the LCIC is introduced is to provide a criterion that satisfied both the consistency requirements and did not produce overly parsimonious estimates, commonly obtained by the BIC (Gonzalo and Pitarakis, 1998). The LCIC performed better when compared to the AIC, BIC and HQ criteria.

Barr and Kantor (2004) use a nonlinear composite index to assess the suitability of one-day batsmen. The new two-dimensional criterion includes the strike rate and batting average of the batsman. These criteria are included due to the nature of one-day cricket. As one-day cricket has limited overs it is essential that batsmen not only focus on their batting average but also on their strike rate (Barr and Kantor, 2004). This new criterion is computed using a nonlinear approach as each criterion used made a proportional contribution to the new criterion instead of an additive contribution. This criterion is found to perform well when assessing the suitability of batsmen for one-day cricket matches.

At the time of this study no composite index had been sourced in the literature within the SEM domain to the researcher's knowledge. Due to the controversy regarding the best method of assessing model fit, the domain may benefit from the use of a composite index. In this study, a composite index is developed to provide a criterion that performs better than the traditional goodness-of-fit indices when assessing model fit. This is done by combining four indices which are chosen based on their frequency of use within the literature and their theoretical properties. The frequency of use is important to consider as it provides information about which indices are commonly understood and available to researchers. The theoretical properties, such as how indices handle conditions such as model misspecification and sample size, are important to ensure that fit is assessed across all conditions. The number of indices included in the CI is based on the suggestions of researchers such as Byrne (2010), Hair et al. (2014) and Kline (2016). The CI is constructed to assess the overall model fit by reviewing the average performance of the indices instead of interpreting each index on its own. The motivation for the CI is that using a combination of several indices will perform better than the traditional method of assessing each individual index. Only one CI was proposed in this research as it encompassed the commonly suggested indices for both types of indices necessary in the assessment of fit and due to the novelty within this domain.

3.3.1 Development

In the development of the composite index, the indices most frequently reported were initially reviewed. The most frequently reported indices' properties were then investigated. The indices included in the CI were chosen based on a combination of its frequency and theoretical properties.

As seen in the frequency of goodness-of-fit indices section, the χ^2 statistic was the most frequently reported measure. This statistic is excluded from the CI due to its limitations and for the CI to be developed as a stand-alone index, however it is suggested that it is reported, along with the CI, when assessing model fit. The next most frequently reported indices are the CFI and TLI. These indices were both included in the CI due to their frequency of use in the literature. The TLI and CFI were also found to favour parsimonious models and were relatively insensitive to sample size (Schreiber et al., 2006). These indices were also found to perform well in the study by Hu and Bentler (1999). Therefore, these indices are frequently used and possess desirable properties. Both indices are incremental indices, measuring how well the estimated covariance matrix compares to a null model's covariance matrix.

The most frequently reported absolute fit index was the RMSEA and is included in the composite index. The RMSEA is found to be relatively insensitive to sample size and favours parsimonious models. The final absolute fit index in the CI is the SRMR which, while not one of the more frequently reported indices, it is included based on its theoretical properties of misspecification sensitivity and insensitivity to sample size (Iacobucci, 2009). These indices also formed part of the Two-Index presentation strategy proposed by Hu and Bentler (1999). These indices are included to assess how well the model fitted the sample data.

Therefore, the indices included in the CI are incremental and absolute indices. No parsimony indices are included in the model as the assumption that researchers will always compare multiple models is not made.

3.3.2 Aggregation and weightings

The CI is developed by linearly aggregating the indices chosen to form the following formula,

$$CI = \frac{1}{4}CFI + \frac{1}{4}TLI + \frac{1}{4}(1 - RMSEA) + \frac{1}{4}(1 - SRMR) \qquad (3.12).$$

Equal weightings are used to develop the CI as there is no theoretical backing indicating that one index should have a larger weighting. The absolute fit indices chosen, RMSEA and SRMR, are reversed within the CI. This is done to keep the directionality, indicating a good model fit, consistent. The reversed indices are noted with an asterixis, RMSEA* and SRMR* from this point onwards. The CI is developed in such a way that a higher value indicates a better model fit. A cut-off point is obtained by reviewing the values that indicate a good model fit for each index. These values are then inserted into the CI formula to obtain a value to indicate a good model fit. This is used as a cut-off point to be used by researchers. This is done for sample sizes of less than and greater than 250, as suggested by Hair et al. (2014).

Table 3.2: Goodness-of-fit index cut-off points for different sample sizes

Index	**N < 250**	**N > 250**
CFI	> 0.95	> 0.90
TLI	> 0.97	> 0.95
RMSEA*	> 0.95	> 0.92
SRMR*	> 0.92	> 0.90
CI	> 0.9475	> 0.9175

The cut-off criteria shown in Table 3.2 are based on the conventional criteria. The indices included in the CI are chosen due to their desirable properties, which are empirically assessed in the following chapters. The indices chosen are also the same as the indices in the Two-Index presentation strategy by Hu and Bentler (1999). This provides added validation for the indices chosen in the CI.

3.4 Conclusion

In closing, a CI is proposed in this study as an alternative method of assessing model fit. The CI is developed based on the reporting frequency of each index in the literature and the properties of each index. The CI is developed using two incremental fit indices and two absolute fit indices, to assess the model's fit relative to a null model and to assess how well the model replicates the data. Both types of indices are important to report as it measures the overall model fit. The incremental fit indices included are the TLI and CFI, and the absolute fit indices included are the RMSEA and SRMR. The indices are linearly aggregated with equal weightings, as no theory is found to suggest that unequal weightings are required. The CI's performance is evaluated using a Monte Carlo simulation study after which a case study is performed.

Chapter 4 The Simulation Methodology

4.1 Introduction

To assess whether the CI provides a better method of model assessment compared to the traditional method of reporting multiple indices, a simulation study is undertaken. The behaviour of the CI is evaluated across a range of sample sizes, two estimation methods and two misspecified models, to compare its performance to the traditional method of assessing model fit. The methodology used to obtain these results is explained in this chapter. The software used in this study is discussed in the following subsection.

4.2 Software

The software options for SEM was discussed in chapter 2. The R statistical environment was chosen to run the simulation study and analyse the models. It was chosen due to its availability and feasibility as R is a free and easily available (Hallgren, 2014). R is also compatible across various operating systems allowing results to be reproducible.

This software has the capability to fit many different models and has several R packages available for modelling SEM (Lee, 2015). Packages have been developed in the same notation as the traditional SEM software such as LISREL, AMOS and MPLUS. The R packages chosen to perform the simulation are OpenMx (Neale et al., 2016) and SimSem (Pornprasertmanit et al., 2016).

4.3 Simulation study

The main research aim of this study is to compare the performance of a composite index to the performance of the traditional method of reporting several goodness-of-fit indices, in the assessment of model fit. The simulation study also investigates the performance of the proposed CI under varying experimental conditions. These conditions include sample size, estimation method and model misspecification. The remaining steps in the simulation process are divided into subsections based on these conditions. Once the simulations are executed, each dataset is briefly assessed to ensure that the correct number of replications are analysed and to ensure the correct indices are captured. This covered the troubleshooting and verification step of the simulation study.

4.3.1 The simulation model

A Monte Carlo simulation study is conducted using a three-factor model (Figure 4.1) with varying sample sizes. The three-factor model (Figure 4.1) used in the simulation is based on a model obtained in a previous case study. The initial simulation model is created using three latent factors (F1, F2, and F3) and 14 observed variables (denoted x1 – x14). The model allows correlational relationships between the latent factors. This model is chosen as it represented a theoretical model which is discussed in further detail in chapter 6. The simulation model is constructed using the OpenMx syntax. An OpenMx model is set up based on the model in Figure 4.1, using the parameter values given in Table 4.1 as starting values for the simulation. The parameter values are the unstandardised estimates as the OpenMx program uses scale identification in the calculation of parameter estimates. This means that the first loading of each latent factor is fixed to one and all factor variances are free. This is the commonly used method of calculating parameter estimates in SEM. An OpenMx model is built using four components. These components include a regression coefficient matrix, covariance matrix, filter matrix and mean matrix.

.

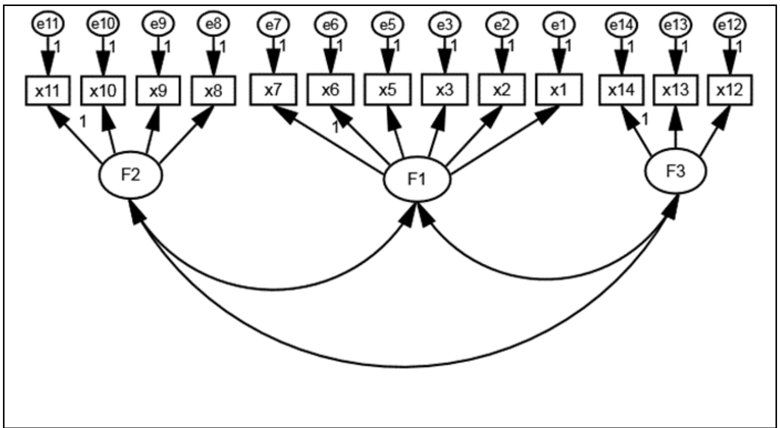


Figure 4.1: Simulation model

The regression coefficient matrix contains two sub-matrices that are used in its calculation. The first sub-matrix represents the unstandardised factor loadings between the latent factors and their observed variables. These values are used as starting values in the simulation process. The second sub-matrix is used to indicate which of the unstandardised factor loadings are fixed and which are freely estimated. The fixed parameters relate to the scale identification method used in the program. These sub-matrices are brought together to define the full regression coefficient matrix.

The covariance matrix is also defined by two sub-matrices. The first sub-matrix contains the error variances associated with each observed variable as well as the covariances between the latent factors. The second sub-matrix indicates whether any variance or covariance pathways are fixed or if each can be freely estimated. All variances and covariances are freely estimated in this simulation as the scale identification method is employed.

The filter matrix represents the number of observed variables and latent factors. The number of rows represents the number of observed variables while the number of columns represents both the number of observed variables and latent factors in the model. The mean matrix is then constructed to contain the mean population values calculated within the simulation process.

Table 4.1: Population parameter estimates

Parameter pathway	**Parameter value**
x1 ← F1	1
x2 ← F1	0.81
x3 ← F1	0.91
x4 ← F1	0.96
x5 ← F1	1.05
x6 ← F1	1.94
x7 ← F1	1.14
x8 ← F2	1
x9 ← F2	0.84
x10 ← F2	1.06
x11 ← F2	1.23
x12 ← F3	1
x13 ← F3	0.75
x14 ← F3	0.96
F1 ↔ F2	0.23
F1 ↔ F3	0.26
F2 ↔ F3	0.29

The model is then built using these four matrices. The simulated data can be calculated using the OpenMx package however this package does not allow for multiple replications to be performed. Therefore, a Monte Carlo simulation can be performed using the SimSem package. This package simulates data based on the OpenMx model built and allows for multiple replications to be performed. Once simulated the data is exported into an Excel file and used to analyse the models and the results for each replication are captured.

4.3.2 Sample size

The first experimental condition investigated is sample size. The following sample sizes are investigated, N = 25, 50, 75, 100, 300, 500, 1000. Each dataset of a specific sample size is replicated 200 times (Hu and Bentler, 1999). The sample size 25 was later removed from the study as it failed to converge. The first simulation consists of 1200 (6 sample sizes x 200 replications) datasets to analyse.

Each dataset of differing sample sizes is analysed per the model in Figure 4.1 and the various goodness-of-fit indices for each model obtained. The indices used in the construction of the CI, namely the CFI, TLI, SRMR, and RMSEA, are extracted and captured for each dataset. The results for the SRMR* and RMSEA* are calculated and captured. From those values, the CI value is calculated for each replication using equation 3.12 from Chapter 3, and the average taken of all the CI values to obtain an overall CI value per sample size. The differences in average value index values is compared across the sample sizes.

To assess the difference in performance of the CI versus the traditional method of reporting model fit, each index result is classified as either a poor, adequate, or a good model fit, according to the respective cut-off values. As stated in the literature, it is suggested practise to report three to four indices to assess model fit (Byrne, 2010; Hair et al., 2014; Kline, 2016). Therefore, the five possible combinations of the CFI, TLI, RMSEA, and SRMR are reviewed and the number of times that all the indices, within each combination, correctly classified the models, are counted. The five combinations investigated are shown in Table 4.2.

Table 4.2: Index combinations compared against CI

Combination number	**Index combinations**
1	CFI, TLI, RMSEA, and SRMR
2	CFI, TLI, and RMSEA
3	CFI, TLI, and SRMR
4	TLI, RMSEA, and SRMR
5	CFI, RMSEA, and SRMR

The performance of each combination and the CI, stated as a percentage, for each sample size is calculated and compared. A one-tailed two population proportions test is performed to determine the statistical significance, at a 5% level, of the performance difference between the combinations and the CI (Stangroom, 2017). The assumptions of the test are assessed prior to the assessment. This test assesses the following hypotheses,

$$\begin{aligned} H_0 &: p_{CI} = p_{combination} \\ H_1 &: p_{CI} > p_{combination} \end{aligned} \qquad (4.1).$$

The following test statistic is used in the completion of the test,

$$Z = \frac{(\bar{p}_1 - \bar{p}_2) - 0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (4.2),$$

where $\overline{p}_1$ is the proportion for population 1 (CI), $\overline{p}_2$ is the proportion for population 2 (Combination), n_1 is the sample size for population 1 (CI), n_2 is the sample size for population 2 (Combination), and $\overline{p}$ is the pooled estimate of the proportions. The pooled estimate of the proportions is defined as follows,

$$\overline{p} = \frac{\overline{p}_1 n_1 + \overline{p}_2 n_2}{n_1 + n_2} \quad (4.3).$$

The results of this test are used to provide statistical support to the differences observed in performance between the CI and index combinations.

4.3.3 Estimation method

The next experimental condition to be investigated is the estimation method. The two estimation methods chosen to be investigated in this study are maximum likelihood (ML) estimation and generalised least squares (GLS) estimation. These estimation methods are used due to their frequency of use within the literature (Fan et al.,1999).

The simulation model (Figure 4.1) is the model used to analyse the performance of the CI under varying estimation methods. The result of each index used in the construction of the CI is captured and recorded for each replication across both estimation methods. Estimation method is investigated to identify whether the CI performed better than the traditional combinations under both estimation methods. A comparison of the performance of the CI across the two estimation methods is also performed.

The CI value for each replication was calculated and averaged over all replications to obtain an overall CI value. The average CI value was compared to the average value for each component within the CI, CFI, TLI, RMSEA, and SRMR.

Each replication of each index, including the CI, is then classified according to the level of model fit. Each combination, found in Table 4.2, is then classified as to whether or not the model is classified correctly. The performance of each combination is then compared to the performance of the CI and tested for statistical significance, using the one-tailed proportions test described in section 4.3.2. This process is repeated for each estimation method and sample size resulting in 2400 results (2 estimation methods x 6 sample sizes x 200 replications).

4.3.4 Model misspecification

The final experimental condition to be investigated within this study is model misspecification. Two modified models are analysed based on the original model with varying levels of misspecification. Misspecification is a subjective topic as there is no objective way to define misspecifications (Fan et al., 1999). The two misspecified models, denoted M1.1 and M1.2, chosen are shown in Figures 4.2 and 4.3 below.

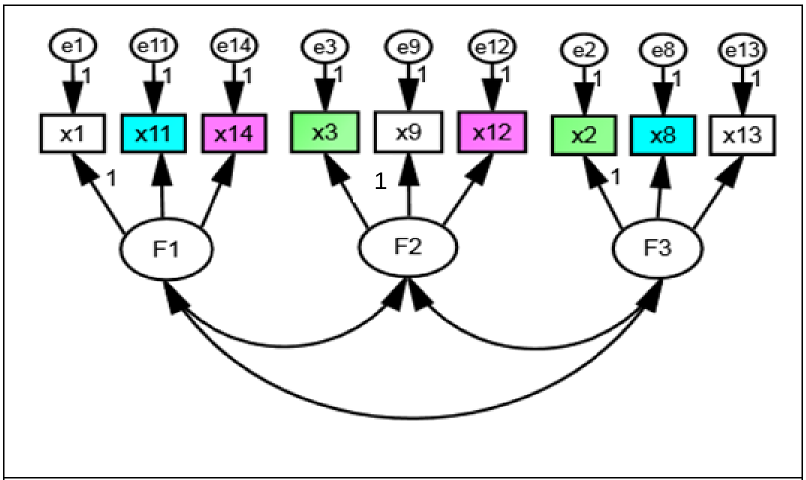


Figure 4.2: Misspecified simulation model 1 (M1.1)

The first misspecified model (Figure 4.2) contains three latent factors, each defined by three observed variables. Each new factor is defined by an item from each one of the factors within the original simulation model. Factor one (F1) is defined by items x1, x11, and x14, factor two (F2) is defined by items x3, x9, and x12, and finally, factor three (F3) is defined by items x2, x8, and x13. The items originally moved from factor one, two and three are represented by the colours lime, aqua and fuchsia, respectively. This model is classified as a medium-misspecification as it contained the same number of factors as the original simulation model, however the items used to define the factors are different.

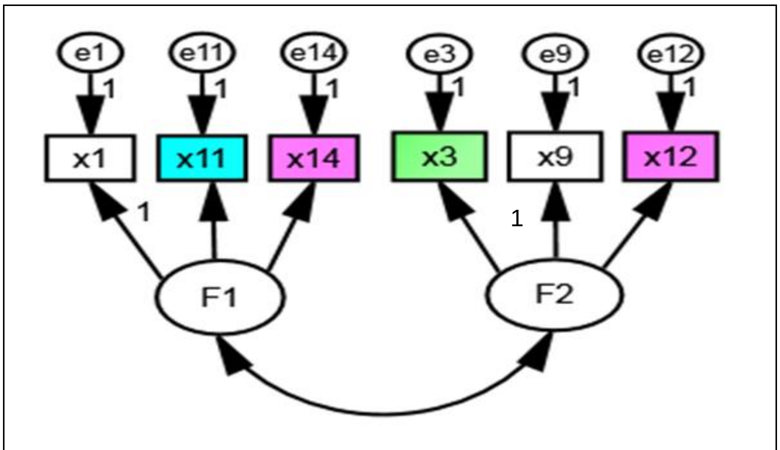


Figure 4.3: Misspecified simulation model 2 (M1.2)

The second misspecified model (Figure 4.3) contains two factors, each defined by three items. Factor one (F1) within this model is defined by items x1, x11, and x14, and factor two (F2) is defined by items x3, x9, and x12. The second misspecification model, is classified as a high-misspecification as it contained two factors compared to the three factors as in the original simulation model, as well as having different items defining the factors.

The two misspecified models are analysed for each sample size and estimation method. The model fit results (CFI, TLI, RMSEA, SRMR, and CI) are captured for each replication of each combination of model, sample size and estimation method. A total of 4800 (6 sample sizes x 2 misspecified models x 2 estimation methods x 200 replications) models are analysed.

The CI is calculated for each replication and averaged across all the replications to obtain an overall CI value for each misspecified model, for each estimation method and each sample size. The average values for each index, CFI, TLI, RMSEA, and SRMR, is calculated and compared to the CI. The index values for each replication is classified into three categories, poor, adequate, and good model fit, according to the respective cut-off values. Following the classification of each index, the combinations, found in Table 4.2, are classified according to whether all the indices within the combination correctly classified the model. The performance of these combinations is captured as a percentage, and compared to the performance of the CI. The performance differences between the combinations and CI are then tested for statistically significant differences to determine whether the CI performed better than reporting multiple indices. This is done by performing a one-tailed population proportions test for each comparison, to determine whether the CI significantly outperforms the multiple combinations. This test is defined in section 4.3.2. Once all the experimental conditions are assessed, the overall performance of the CI is discussed.

4.4 Conclusion

The CI proposed in chapter three is tested using a Monte Carlo simulation study to assess its behaviour under varying experimental conditions. The conditions tested are sample size, estimation methods and level of misspecification. Each unique model is run and the model fit assessed by reviewing the indices used in the construction of the CI. Each index is classified as to whether it correctly classifies the model. These classifications are used to determine the performance of the index combinations, and then compared to the performance of the CI results. The performance differences are then tested for statistical significance to draw conclusions regarding whether the CI outperforms the traditional method of assessing model fit, the objective of this study.

Chapter 5 Results and Discussion

5.1 Introduction

The CI developed in chapter three was assessed using a simulation study to examine its benefits under varying experimental conditions by comparing the performance of the CI compared to combinations of the traditional indices. The experimental conditions investigated were sample size, estimation method and model misspecification. These results, obtained using the methodology described in the previous chapter, are discussed in the following sections, sample size, estimation method, and model misspecification.

5.2 Sample size

The simulation study was conducted for a three-factor model with varying sample sizes. The following results involve the sample sizes N = 50, 75, 100, 300, 500, 1000. The results of each replication were captured per sample size and the necessary values extracted. Table 5.1 represents an example of the goodness-of-fit index results of one replication for sample size 50. The example of the parameter estimates is found in appendix A. Once all the replication results had been captured, the values for the CFI, TLI, RMSEA and SRMR are extracted. Using these values, the CI is calculated. Using this example, the index results are 0.933, 0.918, 0.069 and 0.086 for the CFI, TLI, RMSEA and SRMR, respectively. The CI result for this example is 0.924.

Table 5.1: Result output example

Number of observations	50	
Estimator	ML	
Chi-square test statistic	91.502	
Degrees of freedom	74	
P-value	0.082	
Incremental fit:		
Comparative Fit Index (CFI)	0.933	
Tucker-Lewis Index (TLI)	0.918	
Root Mean Square Error of Approximation:		
RMSEA	0.069	
90 percent Confidence Interval	0.000	0.111
P-value RMSEA	0.261	
Standardised Root Mean Square Residual:		
SRMR	0.086	

Using the results of each replication, the results for each index used in the CI, as well as the CI itself, were collected and averaged and were summarised in Table 5.2. The distribution of the CI values calculated from each replication for sample size 50 is represented in Figure 5.1. An adequate model fit is indicated by values greater than 0.9175 but less than 0.945. A good model fit is indicated by values greater than 0.945. For this model, most of the CI values should be within these ranges.

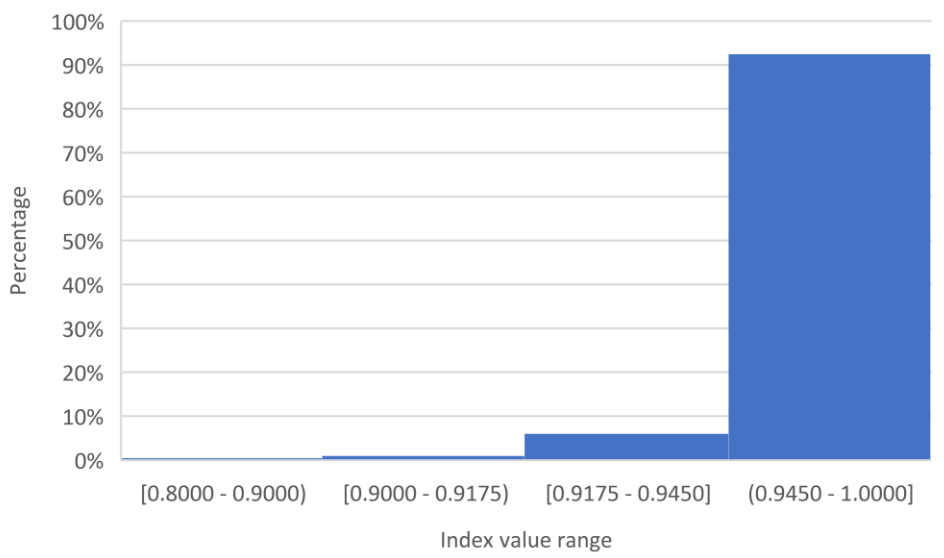


Figure 5.1: Histogram of CI results for sample size 100

From Figure 5.1 it is observed that 93% of the CI results fall in the range of 0.945 to 1.00, indicating a good model fit, and 6% of the results fall in the range of 0.9175 to 0.945, indicating an adequate fit. Therefore, a total of 99% of results fall in the adequate to good fit range. This is a desirable result. The CI results for sample sizes 50 and 75 range from 73% to 97.5% falling within the adequate to good fit criteria. This indicates that the CI results may be influenced by sample size.

Once the results have been captured, the average value of each index used in the CI construction and the CI itself is calculated. These average values are summarised in Table 5.2. Each value within the table is obtained using the general equation,

$$\bar{x} = \frac{1}{200}\sum_{i=1}^{200} x_i^{index} \quad (5.1).$$

If the CFI is used as an example, the average value is calculated as follows,

$$\bar{x}^{CFI} = \frac{1}{200}(0.932 + \cdots + 0.985) = 0.951 \quad (5.2).$$

From Table 5.2 it is observed that as sample size increases, so too does the average index value. Ideally, each index value should be consistent across sample sizes, as each simulation was performed on the same model. By comparing the index values per sample size to their respective cut-off values, the average index values for the TLI, RMSEA and CI at sample size 50, indicate an adequate model as the values are below 0.970 and 0.945 for the TLI and CI, and above 0.05 for the RMSEA. These cut-off values are used to indicate a good model fit. The remaining average values indicated a good model fit compared to their respective cut-off values discussed in chapter three. The cut-off criteria used to indicate a good model fit for the CFI and SRMR are 0.95 and 0.08, respectively.

Table 5.2: Index mean values for different sample size

Index	50	75	100	300	500	1000
CFI	0.951	0.975	0.985	0.997	0.998	0.999
TLI	0.940	0.970	0.981	0.996	0.998	0.999
RMSEA	0.052	0.033	0.025	0.010	0.007	0.005
SRMR	0.078	0.063	0.055	0.031	0.024	0.017
RMSEA*	0.948	0.967	0.975	0.990	0.993	0.995
SRMR*	0.922	0.937	0.945	0.969	0.976	0.983
CI	0.940	0.962	0.972	0.988	0.991	0.994

Once the average index values are obtained for each sample size, the performance of each index is assessed. This is assessed by calculating the number of correct classifications for each index. The results are summarised into three categories defined as poor, adequate or good. The cut-off values associated with these classifications are obtained from the cut-off values discussed in chapter three. An example of these values is found in Table 5.3. Using these cut-off values each replication is classified accordingly. The number of poor, adequate and good classifications are counted and the percentage for each reported in Table 5.4.

Table 5.3: Classification cut-off values

	Classifications		
Index	**Poor**	**Adequate**	**Good**
CFI	< 0.900	> 0.900	> 0.950
TLI	< 0.950	> 0.950	> 0.970
RMSEA	> 0.080	< 0.080	< 0.050
SRMR	> 0.100	< 0.100	< 0.080
RMSEA*	< 0.920	> 0.920	> 0.950
SRMR*	< 0.900	> 0.900	> 0.920
CI	< 0.9175	> 0.9175	> 0.945

The performance of the CI was compared to the performance of the five combinations of the four indices that were used to construct the CI. The performance of the CI is calculated by classifying the number of times all the indices within the combinations classified the model correctly. Should one index incorrectly classify the model, the combination will show an incorrect classification. This is repeated for each replication and the number of correct classifications counted and the percentage calculated. The comparison of the performance of the CI to the index combinations is done to assess whether the CI outperforms the performance of the multiple indices reporting method.

Table 5.4: Correct classification for each index

Index	50	75	100	300	500	1000
CFI	57.00%	82.00%	94.00%	100.00%	100.00%	100.00%
TLI	39.00%	56.00%	74.00%	100.00%	100.00%	100.00%
RMSEA	42.00%	66.50%	85.50%	100.00%	100.00%	100.00%
SRMR	58.50%	98.00%	100.00%	100.00%	100.00%	100.00%

Figure 5.2 shows the level of correct classification (performance) for each combination and the CI. For the sample sizes 50, 75, and 100, the CI performs better in comparison to the different combinations. It is seen that there is an increase in the performance of both combinations and CI as the sample size increases.

For sample sizes 300, 500, and 1000, all combinations and the CI showed a perfect performance thereby indicating for the large samples, any method can be used. The results indicate that the CI outperforms the index combinations at small sample sizes, while it matches the performance of the combinations at larger sample sizes.

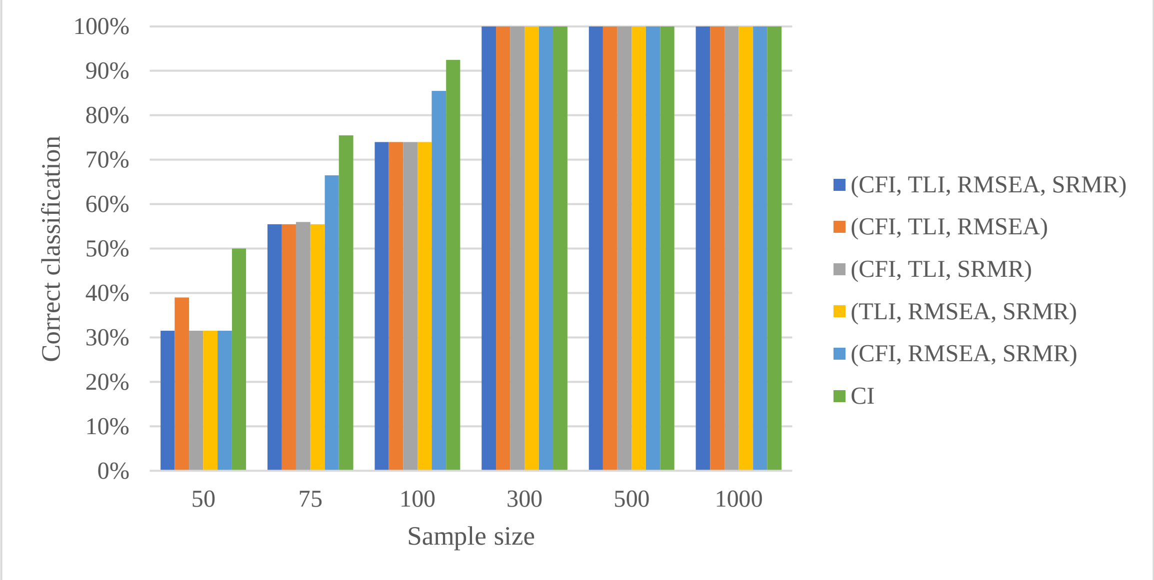


Figure 5.2: Performance of CI in relation to the five combinations

The differences in performance between the CI and the index combinations are tested inferentially using a one-tailed proportions test. The reason for this test is to determine whether the proportion (or performance) of the CI is greater than that of the respective combinations. Prior to performing this test, the assumptions were considered. Since the data is categorical in the form of a binary variable, indicating correct classification or incorrect classification, and there are 200 replications per condition, the assumption of normal approximation is satisfied. Therefore, the test is performed.

Table 5.5: Proportion tests for sample size 50

Comparison (% correct)	P-value	Outcome
CI (50.00%) Combination 1, 3, 4 & 5 (31.50%)	0.00008	Significant
CI (50.00%) Combination 2 (39.00%)	0.01355	Significant

For sample size 50, the CI correctly classified 50% of the models while combinations 1, 3, 4 and 5 classified 31.5% of the models correctly and combination 2 correctly classified 39% of the models. Despite the index averages in Table 5.2 showing that overall the indices represented a good model fit for sample size 50, the performance of the combinations reflects a low percentage. This is because for a combination to correctly classify a model, all indices within the combination must reflect this result. The results of the classification percentages for each index can be found in appendix B.

The performance differences between the CI and the combinations are tested and found to be statistically significant as observed in Table 5.5 (refer to Table 4.2 for index combinations). This indicates that the CI, statistically, performs better than the combinations.

Table 5.6: Proportion tests for sample size 75

Comparison (% correct)	P-value	Outcome
CI (75.50%) Combination 1, 2 & 4 (55.50%)	0.00000	Significant
CI (75.50%) Combination 3 (56.00%)	0.00000	Significant
CI (75.50%) Combination 5 (66.50%)	0.02385	Significant

Table 5.6 contains the results of the proportion tests for sample size 75. The CI correctly classified 75.5% of the models while combinations 1, 2 and 4 classified 55.5% of the models correctly. Combinations 3 and 5 correctly classified 56% and 66.5% of the models, respectively. The differences in performance are tested and it is observed that the CI significantly outperforms all the combinations.

Table 5.7 contains the results of the proportion tests for sample size 100. The CI classified 92.5% of the models correctly. Combinations 1, 2, 3 and 4 classified 74% of the models correctly while combination 5 correctly classified 85.5% of the models. From the results observed in Table 5.7 the CI significantly outperforms all the combinations at a 5% significance level.

Table 5.7: Proportion tests for sample size 100

Comparison (% correct)	P-value	Outcome
CI (92.50%) Combination 1, 2, 3 & 4 (74.00%)	0.00000	Significant
CI (92.50%) Combination 5 (85.50%)	0.01255	Significant

For sample sizes 300, 500, and 1000, all the combinations and the CI correctly classified 100% of the models therefore no differences in performance are observed between the CI and the combinations.

5.3 Estimation method

The next series of comparisons between the CI and the index combinations considered the influence of different estimation methods. The results of the two estimation methods are summarised in this section. The two estimation methods assessed are ML and GLS estimation. Ideally the estimation method chosen should not have a significant effect on the index values.

Table 5.8 provides an example of the goodness-of-fit results for a replication using ML and GLS estimation. The index values used in the calculation of the CI are in red. The CFI and TLI show the largest difference across the estimation methods. Upon review of the indices CFI, TLI, RMSEA and SRMR, the difference in result is observed for the estimation methods, suggesting that estimation method does affect the indices. Using these examples, the associated CI results are 0.888 and 0.961 for GLS and ML estimation, respectively. Therefore, in this example the GLS method resulted in a bad model fit and the ML method resulted in a good model fit. The histogram summary of the CI results for both estimation methods can be found in Figure 5.3. For both estimation methods, most of the CI values should exceed the value 0.9175.

Table 5.8: Result output example for GLS estimation

Number of observations	50		50	
Estimator	GLS		ML	
Chi-square test statistic	77.245		78.721	
Degrees of freedom	74		74	
P-value	0.375		0.332	
Incremental fit:				
Comparative Fit Index (CFI)	0.877		0.984	
Tucker-Lewis Index (TLI)	0.849		0.980	
Root Mean Square Error of Approximation:				
RMSEA	0.030		0.036	
90 percent Confidence Interval	0.000	0.088	0.000	0.090
P-value RMSEA	0.650		0.612	
Standardised Root Mean Square Residual:				
SRMR	0.146		0.085	

From the histogram of the CI using the ML estimator, 93% of the results fell in the range indicating a good fit, while 6% of the results indicate an adequate fit. The range of the CI for the GLS estimation showed a greater dispersion than the ML method, with CI results ranging from 0.70 to 1.00. Approximately 74% of the results fell in the range that indicates a good fit, while 9% of the results fell in the range indicating an adequate fit. Therefore, 99% and 83% of the CI results indicated either an adequate or good fit, using ML and GLS estimation, respectively. This indicates that the estimation method used may affect the CI results.

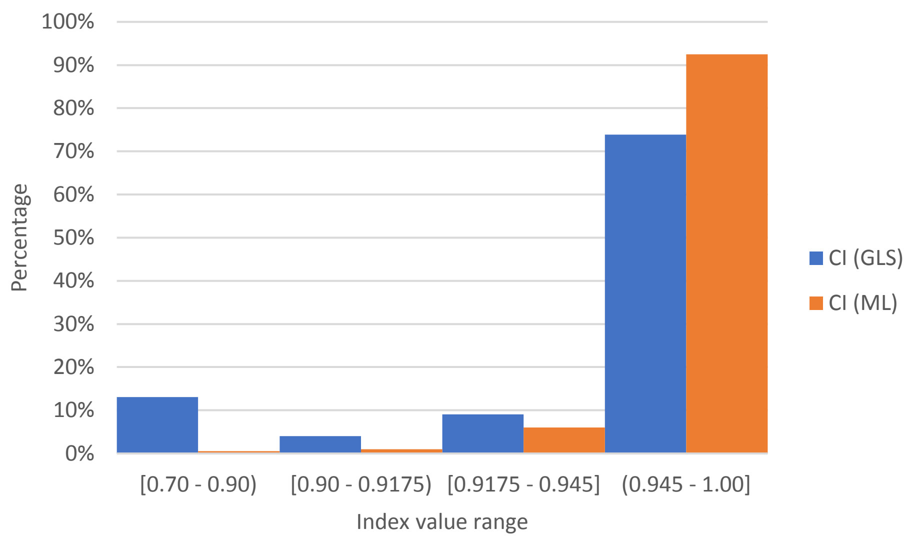


Figure 5.3:Histogram of CI results for ML and GLS estimation at sample size 100

Table 5.9 summarises the average index values per estimation method for the different sample size. As seen from the results of the sample size, as the sample size increases, the average index value increases for both estimation methods. It is also seen that the average value for the RMSEA showed smaller values using the GLS estimation method when compared to the ML estimation method. The opposite relationship is observed for the SRMR average values. The SRMR average values at sample sizes of 50 and 75 indicated a larger discrepancy between estimation methods, ML and GLS, compared to the remaining sample sizes. The CI average values indicate slightly lower values using GLS estimation methods, compared to the ML estimation methods. This suggests that the CI results may be slightly lower if GLS estimation is used.

Table 5.9: Average index values for the two estimation methods.

Index	**50**		**75**		**100**		**300**		**500**		**1000**	
	ML	GLS	ML	GLS	ML	GLS	ML	GLS	ML	GLS	ML	GLS
CFI	0.951	0.958	0.975	0.949	0.985	0.959	0.997	0.983	0.998	0.990	0.999	0.994
TLI	0.940	0.949	0.970	0.937	0.981	0.950	0.996	0.979	0.998	0.987	0.999	0.993
RMSEA	0.052	0.008	0.033	0.012	0.025	0.011	0.010	0.008	0.007	0.006	0.005	0.005
SRMR	0.078	0.189	0.063	0.109	0.055	0.082	0.031	0.037	0.024	0.027	0.017	0.018
RMSEA*	0.948	0.992	0.967	0.988	0.975	0.989	0.990	0.992	0.993	0.994	0.995	0.995
SRMR*	0.922	0.811	0.937	0.891	0.945	0.918	0.969	0.963	0.976	0.973	0.983	0.982
CI	0.940	0.927	0.962	0.942	0.972	0.954	0.988	0.979	0.991	0.986	0.994	0.991

The performance of the CI for each estimation method, relative to the combinations of indices is investigated next. These results can be found in Figure 5.4. An interesting observation is observed in the comparison of the estimation methods' index combinations for sample size 50 (Figure 5.4 (a)). For the GLS estimation method, only combination 2 and the CI, correctly classified the models to some extent. This result was the only instance in which the CI did not outperform or equal the performance of the combinations within this study. This was because the SRMR failed to correctly classify any of the models. This, therefore, had an impact on the classification of any combination containing the SRMR. The results of the individual index classifications are summarised in Table 5.10.

Table 5.10: Correct classification for both estimation methods

Index	Estimation	50	75	100	300	500	1000
CFI	ML	57.00%	82.00%	94.00%	100%	100%	100%
	GLS	79.01%	69.50%	73.87%	85.95%	96.31%	99.33%
TLI	ML	39.00%	56.00%	74.00%	100%	100%	100%
	GLS	75.93%	66.00%	69.35%	70.90%	78.86%	95.33%
RMSEA	ML	42.00%	66.50%	85.50%	100%	100%	100%
	GLS	96.30%	94.50%	94.97%	100%	100%	100%
SRMR	ML	58.50%	98.00%	100%	100%	100%	100%
	GLS	0.00%	7.50%	48.24%	100%	100%	100%

The differences in performance are assessed using a one-tail proportions test as indicated previously (Table 5.11). The CI correctly classified 62.35% of the models using the GLS estimation. Combination 2 correctly classified 75.93% of the models. The difference between the CI and combination 2 is shown to be statistically significant favouring combination 2. The statistical differences for ML estimation were examined in the previous subsection, where the CI was shown to significantly outperform all the combinations under this estimation method.

Table 5.11: Proportion tests results for GLS estimation for sample size 50

Comparison (% correct)	P-value	Outcome
CI (62.35%)	0.00000	Significant
Combination 1, 3, 4 & 5 (0.00%)		
CI (62.35%)	0.01321	Significant
Combination 2 (75.93%)		

For sample size 75 (Figure 5.4(b)), overall the combinations performed better under ML estimation method, with the combination 5 correctly classifying 66.5% of the models with the CI outperforming the combinations by correctly classifying 75.5% of the models. These differences were assessed in the previous subsection and the CI was shown to statistically outperform the combinations at a 5% significance level.

Table 5.12: Proportion tests results for GLS estimation for sample size 75

Comparison (% correct)	P-value	Outcome
CI (68.00%) Combination 1, 3, 4 & 5 (7.50%)	0.00000	Significant
CI (68.00%) Combination 2 (66.00%)	0.33360	Insignificant

Under the GLS estimation method, at sample size 75, the summary results in Table 5.12 report that combination 2 correctly classified 66% of the models with the CI correctly classifying 68% of the models. The remaining combinations correctly classified 7.5% of the models. The difference in performance between the CI and combination 2 is tested and showed that the difference was statistically insignificant. However, the difference between the CI and the remaining combinations is significant, indicating that the CI outperformed combinations 1, 3, 4, and 5, while the difference was insignificant for combination 2. This indicates that the SRMR does not perform adequately using GLS estimation, as the combinations 1, 3, 4 and 5 contain this index. The combination that does not include SRMR performed in such a way that the difference between the CI and this combination was insignificant.

With a sample size of 100 (Figure 5.4(c)), the performance of the combinations and CI improves in comparison to the previous two sample sizes. Under ML estimation, the CI performs best by correctly classifying 92.5% of the models, with combination 5 correctly classifying 85.5% of the models. The remaining combinations correctly classified 74% of the models. These differences in performance were assessed for statistical significance in the previous subsection and the CI was shown to statistically outperform all the combinations at a 5% significance level.

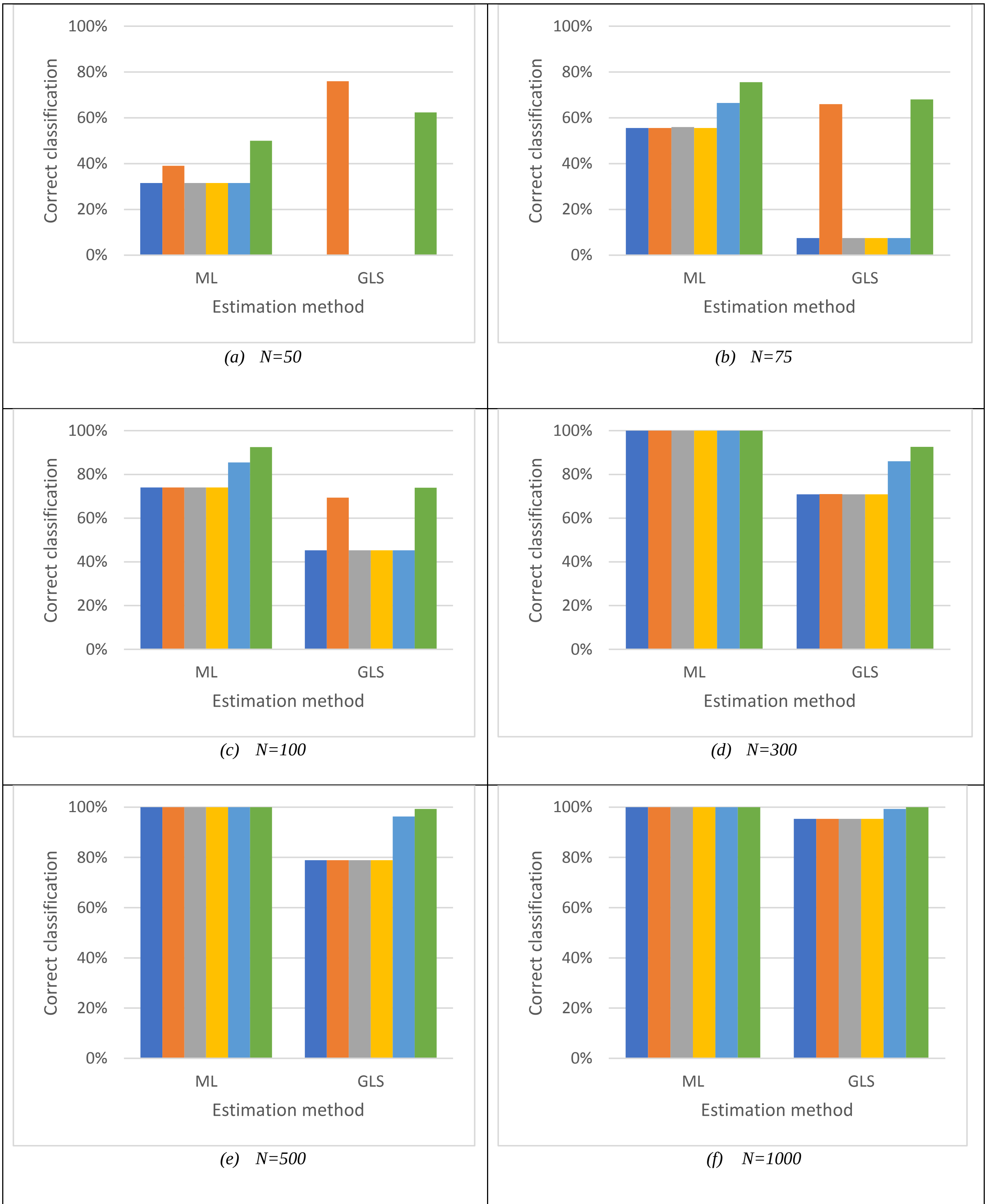


Figure 5.4: Performance of combinations versus CI

(CFI, TLI, RMSEA*, SRMR*) (CFI, TLI, RMSEA*) (CFI, TLI, SRMR*) (TLI, RMSEA*, SRMR*) (CFI, RMSEA*, SRMR*) CI

Table 5.13: Proportion tests results for GLS estimation for sample size 100

Comparison (% correct)	P-value	Outcome
CI (73.87%) Combination 1, 3, 4 & 5 (45.23%)	0.00000	Significant
CI (73.87%) Combination 2 (69.35%)	0.16109	Insignificant

Under GLS estimation, the CI performed best by correctly classifying 73.87% of the models, with combination 2 correctly classifying 69.35% of the models (Table 5.13). The remaining combinations classified 45.23% of the models correctly. The difference in performance of the CI and combination 2 is tested and found to be insignificant. The difference between the performance of the CI and combinations 1, 3, 4, and 5, are compared and found to be statistically significant. Therefore, under GLS estimation, the CI outperforms combinations 1, 3, 4, and 5 significantly, while the difference between the CI and combination 2 is insignificant.

For sample sizes 300, 500, and 1000 (Figures 5.4(d), 5.4(e), and 5.4(f), respectively), under ML estimation, all the combinations and CI correctly classified 100% of all models. For sample size 300, under GLS estimation, the CI performs best by classifying 94.64% of the models correctly. Combination 5 was the second-best performer classifying 85.95% of models correctly, with the remaining combinations correctly classifying 70.9% of the models. The difference between the CI and combination 5 was assessed and found to be statistically significant (Table 5.14). The performance of the CI is compared to the performance of combinations 1, 2, 3, and 4, and found to be significant. Therefore, under GLS estimation, at sample size 300, the CI outperformed all the combinations. This indicates that using the CI would be the more beneficial method of assessing model fit.

Table 5.14: Proportion tests results for GLS estimation for sample size 300

Comparison (% correct)	P-value	Outcome
CI (94.64%) Combination 1, 2, 3 & 4 (70.90%)	0.00000	Significant
CI (94.64%) Combination 5 (85.95%)	0.00402	Significant

For sample size 500, under GLS estimation, the CI correctly classified 99.33% of the models, with combination 5 classifying 96.31% correctly, and the remaining combinations correctly classified 78.86% of the models. The difference in performance of the CI and combination 5 is assessed and shown to be significant (Table 5.15). The performance of the CI and combinations 1, 2, 3, and 4, are assessed and shown to be significant. Therefore, under GLS estimation, at sample size 500, the CI statistically outperforms all the combinations.

Table 5.15: Proportion tests results for GLS estimation for sample size 500

Comparison (% correct)	P-value	Outcome
CI (99.33%) Combination 1, 2, 3 & 4 (78.86%)	0.00000	Significant
CI (99.33%) Combination 5 (96.31%)	0.00587	Significant

For sample size 1000, under GLS estimation, the CI correctly classified 100% of the models, while combination 5 classified 99.33% of the models correctly. The remaining combinations classified 95.33% of the models correctly. The difference between CI and combination 5 is assessed and found to be insignificant at a 5% significance level but was found to be significant at a 10% significance level (Table 5.16). The difference between the CI and combinations 1, 2, 3, and 4, is assessed and found to be significant. Therefore, under GLS estimation, at sample size 1000, the CI outperforms the combinations to some significant extent.

Table 5.16: Proportion tests results for GLS estimation for sample size 1000

Comparison (% correct)	P-value	Outcome
CI (100.00%) Combination 1, 2, 3 & 4 (95.33%)	0.00008	Significant
CI (100.00%) Combination 5 (99.33%)	0.0778	Insignificant

When comparing the performance of the combinations and the CI under different estimation methods, the performance overall was better under the ML estimation method. The same increasing pattern was observed between the sample size and average index values. Overall the GLS estimation method obtained lower levels of performance of the combinations and CI. An interesting pattern was observed, under the GLS estimation method, at smaller sample sizes combination 2 was the second-best performing combination. This changed to the combination 5 in the larger sample sizes. This was because SRMR did not perform adequately at smaller sample sizes. These results indicate that the CI provides a better method of assessing model fit when compared to the traditional methods.

5.4 Model misspecification

The final characteristic investigated using the simulation study was how the indices performed when models were misspecified. Two misspecified models (M1.1 and M1.2) were run using ML estimation and GLS estimation. The results of the two misspecified models are summarised in this section. Table 5.17 and 5.18 contain an example of the results of a replication for the misspecified models under ML and GLS estimation, respectively. The indices used in the calculation of the CI are in red. Upon review of the results in red, the results obtained using ML showed higher index values when compared to the results obtained using GLS estimation. All the results using both estimation methods are below the cut off criteria indicating an adequate model fit, a desirable result for misspecified models. The indices using GLS estimation appear to represent the misspecified models better when compared to the ML estimation, due to the lower values obtained.

Table 5.17: Result output example for misspecified models using ML estimation

	M1.1		M1.2	
Number of observations	50		50	
Estimator	ML		ML	
Chi-square test statistic	45.497		13.433	
Degrees of freedom	24		8	
P-value	0.005		0.098	
Incremental fit:				
Comparative Fit Index (CFI)	0.762		0.841	
Tucker-Lewis Index (TLI)	0.643		0.702	
Root Mean Square Error of Approximation:				
RMSEA	0.134		0.117	
90 percent Confidence Interval	0.072	0.193	0.000	0.222
P-value RMSEA	0.019		0.153	
Standardised Root Mean Square Residual:				
SRMR	0.101		0.092	

Table 5.18: Result output example for misspecified models using GLS estimation

	M1.1		M1.2	
Number of observations	50		50	
Estimator	GLS		GLS	
Chi-square test statistic	38.542		18.791	
Degrees of freedom	24		8	
P-value	0.030		0.016	
Incremental fit:				
Comparative Fit Index (CFI)	0.416		0.565	
Tucker-Lewis Index (TLI)	0.125		0.183	
Root Mean Square Error of Approximation:				
RMSEA	0.111		0.166	
90 percent Confidence Interval	0.035	0.174	0.068	0.265
P-value RMSEA	0.078		0.032	
Standardised Root Mean Square Residual:				
SRMR	0.199		0.150	

The histogram of CI values for the misspecified models under each estimation method is shown in Figures 5.5 and 5.6. A desirable distribution should show that the majority of the results are below the cut-off value indicating a good or adequate model fit. For misspecified model M1.1, 93% of the results using ML estimation indicated a poor fit as these values fell below the cut off criteria. Using GLS estimation method, 68% of the results indicated a poor model fit. This indicates that the CI results differed based on the estimation method.

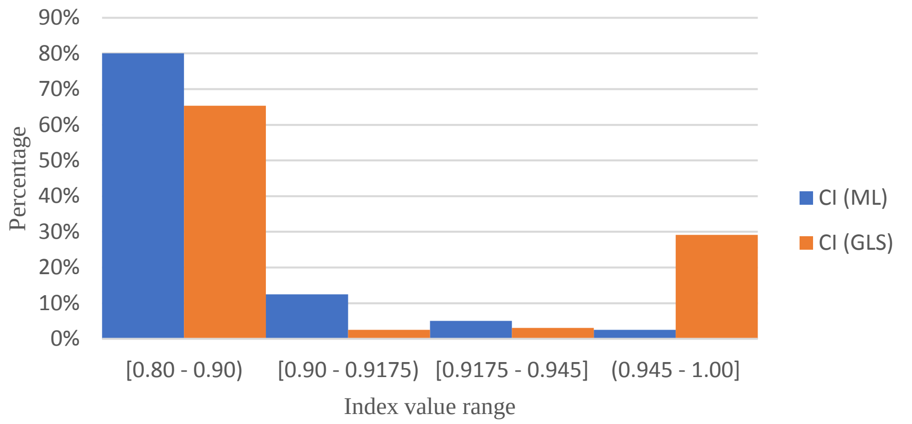


Figure 5.5: CI results for M1.1 using ML and GLS estimation at sample size 100

For misspecified model M1.2, 91% of the results using ML estimation indicated a poor fit as these values fell below the cut off criteria. Using GLS estimation method, 79% of the results indicated a poor model fit. This indicates that the CI results differed based on the estimation method however the GLS estimator results indicated a better result when compared to the previous misspecified model. It is interesting to note that under GLS estimation a proportion of the results indicated a good model fit for misspecified models. This requires further investigation.

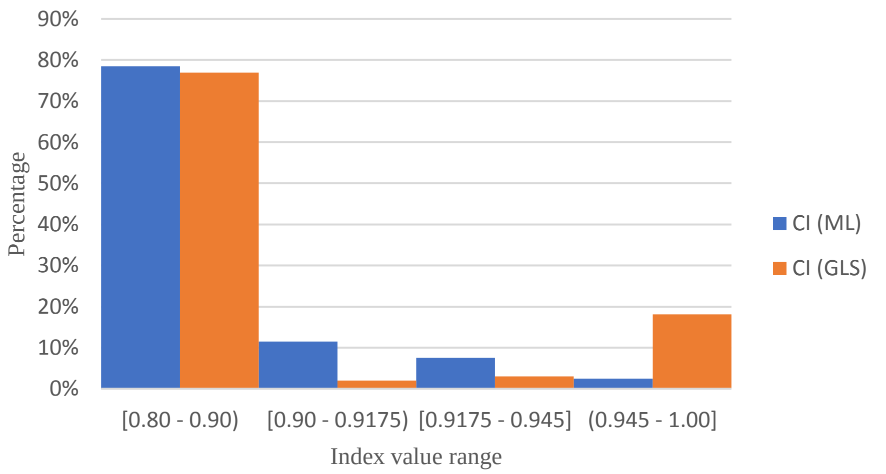


Figure 5.6: CI results for M1.2 using ML and GLS estimation at sample size 100

Table 5.19 contains the average values for each index per misspecified model, under ML estimation. As the index values are calculated for misspecified models, a value below (or above, for certain indices) the cut-off value is expected. This is true for the CFI, TLI, RMSEA, and CI average values, over all the sample sizes. The SRMR average values for sample sizes 50 and 75 fall just above the traditional cut-off value indicating an adequate or bad model fit. The remaining average values fall below the cut-off values, which indicates a good model fit. This indicates that SRMR does not adequately detect model misspecification over the sample sizes which stands in contrast to the literature on this index.

Table 5.19: Average index values for misspecified models (ML Estimation)

Index	**50**		**75**		**100**		**300**		**500**		**1000**	
	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2
CFI	0.861	0.888	0.858	0.888	0.858	0.862	0.862	0.890	0.862	0.892	0.863	0.892
TLI	0.794	0.792	0.788	0.790	0.787	0.791	0.794	0.794	0.794	0.798	0.794	0.798
RMSEA	0.114	0.124	0.118	0.130	0.121	0.121	0.121	0.139	0.122	0.138	0.122	0.139
SRMR	0.093	0.082	0.085	0.074	0.080	0.079	0.069	0.062	0.067	0.060	0.065	0.059
RMSEA*	0.886	0.876	0.882	0.870	0.879	0.879	0.879	0.861	0.878	0.862	0.878	0.861
SRMR*	0.907	0.918	0.915	0.926	0.920	0.921	0.931	0.938	0.933	0.940	0.935	0.941
CI	0.862	0.869	0.861	0.869	0.861	0.863	0.866	0.871	0.867	0.873	0.868	0.873

Table 5.20 contains the average index values for the two misspecified models (M1.1 and M1.2) under GLS estimation. Once again, as the index values are calculated for misspecified models, it is expected that the average values should be below (or above) the cut-off value that indicates a good model fit. This pattern is observed for the indices CFI, TLI, SRMR and CI. The TLI average values calculated are especially low, with the highest value reaching 0.688, far below the cut-off value of 0.95 indicating that the TLI detects misspecified models well. Under GLS estimation, the RMSEA did not detect misspecification to the same extent as the other indices, as most of the average values are close to, or below, the cut-off value of 0.08 and 0.05 that indicate an adequate and good model fit, respectively. The classification percentages for the individual indices are summarised in Table 5.21.

Table 5.20: Average index values for misspecified models (Generalised Least Squares)

Index	**50**		**75**		**100**		**300**		**500**		**1000**	
	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2	M1.1	M1.2
CFI	0.718	0.766	0.713	0.726	0.708	0.739	0.755	0.702	0.765	0.680	0.781	0.693
TLI	0.597	0.592	0.590	0.516	0.572	0.530	0.645	0.478	0.659	0.453	0.688	0.477
RMSEA	0.053	0.080	0.057	0.098	0.062	0.099	0.055	0.111	0.052	0.116	0.047	0.110
SRMR	0.219	0.149	0.164	0.133	0.147	0.120	0.106	0.117	0.102	0.124	0.098	0.122
RMSEA*	0.947	0.920	0.943	0.902	0.938	0.901	0.945	0.889	0.948	0.884	0.953	0.890
SRMR*	0.781	0.851	0.836	0.867	0.853	0.880	0.894	0.883	0.898	0.876	0.902	0.878
CI	0.761	0.782	0.771	0.753	0.768	0.763	0.810	0.738	0.817	0.723	0.831	0.734

Table 5.21: Classification percentages for individual indices

Index	**Model**	**Estimation**	**50**	**75**	**100**	**300**	**500**	**1000**
CFI	M1.1	ML	64.00%	70.50%	75.00%	92.33%	97.33%	99.33%
	M1.2		46.50%	48.50%	73.00%	60.00%	62.33%	69.00%
	M1.1	GLS	64.57%	66.00%	67.34%	47.83%	43.29%	37.33%
	M1.2		64.04%	77.00%	76.88%	74.25%	75.84%	69.33%
TLI	M1.1	ML	87.00%	94.50%	98.00%	99.67%	100%	100%
	M1.2		82.00%	88.50%	98.00%	100%	100%	100%
	M1.1	GLS	68.57%	69.50%	70.85%	59.20%	48.99%	37.67%
	M1.2		72.47%	81.50%	82.41%	79.93%	78.52%	69.33%
RMSEA	M1.1	ML	77.50%	88.00%	95.50%	99.67%	99.33%	100%
	M1.2		76.00%	83.50%	95.50%	99.67%	99.33%	100%
	M1.1	GLS	29.71%	39.50%	42.71%	45.48%	42.28%	37.33%
	M1.2		50.00%	66.50%	66.83%	73.24%	75.50%	69.33%
SRMR	M1.1	ML	31.50%	17.00%	12.50%	0.00%	0.00%	0.00%
	M1.2		17.00%	9.50%	10.50%	0.00%	0.00%	0.00%
	M1.1	GLS	97.71%	72.50%	52.26%	23.41%	22.48%	23.33%
	M1.2		75.00%	58.50%	39.70%	28.43%	30.20%	32.33%

Figure 5.7(a) represents the performance of each combination and CI across each misspecified model per estimation method at sample size 50. Across all models, the CI performed best by correctly classifying 78%, 67%, 68%, and 68.54% the misspecified models M1.1 and M1.2 for ML estimation and GLS estimation, respectively. The second-best performing combination for both misspecified models under ML estimation was combination 2, while under GLS estimation combination 3 performed second-best for both misspecified models.

Table 5.22: Proportion test results for sample size 50

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (78.00%) Combination 1, 3, 4 & 5 (31.50%)	M1.1	ML	0.00000	Significant
CI (78.00%) Combination 2 (63.50%)			0.00071	Significant
CI (67.00%) Combination 1, 3, 4 & 5 (17.00%)	M1.2		0.00000	Significant
CI (67.00%) Combination 2 (46.50%)			0.00000	Significant
CI (68.00%) Combination 1, 2, 4 & 5 (29.71%)	M1.1	GLS	0.00000	Significant
CI (68.00%) Combination 3 (64.57%)			0.24825	Insignificant
CI (68.54%) Combination 1, 4 & 5 (47.19%)	M1.2		0.00000	Significant
CI (68.54%) Combination 2 (50.00%)			0.00019	Significant
CI (68.54%) Combination 3 (54.49%)			0.00326	Significant

Combinations 1, 3, 4 and 5 correctly classified 31.5% of the models while combination 2 correctly classified 63.5% of the models. The difference in performance between CI and combinations 1, 3, 4, and 5, as well as combination 2, for M1.1 (ML) are assessed and shown to be statistically significant (Table 5.22).

For model M1.2 (ML) combinations 1, 3, 4 and 5 correctly classified 17% of the model while combination 2 classified 46.5% of the models correctly. This discrepancy can be associated with the presence of the SRMR within the combinations as the SRMR did not adequately classify the misspecified models. The performance of the CI and combinations for M1.2 (ML) is assessed and the CI is shown to significantly outperform all the combinations (Table 5.22).

For model M1.1 under GLS estimation, combinations 1, 2, 4 and 5 correctly classified 29.71% of the models and combination 3 classified 64.57% of the models correctly. The CI is shown to significantly outperform combinations 1, 2, 4, and 5 (Table 5.22), while the difference in performance between the CI and combination 3 was found to be insignificant.

For M1.2 (GLS), combinations 1, 4 and 5 correctly classified 47.19% of the models, combination 2 classified 50% correctly and combination 3 classified 54.49% correctly. The performance of the CI is compared to all combinations and the difference found to be significant (Table 5.22). Overall, at sample size 50, the CI significantly outperformed all the combinations across the misspecification models except for combination 3 in M1.1 (GLS).

Figure 5.7: Performance across misspecified models

(CFI, TLI, RMSEA*, SRMR*) (CFI, TLI, RMSEA*) (CFI, TLI, SRMR*) (TLI, RMSEA*, SRMR*) (CFI, RMSEA*, SRMR*) CI

Table 5.23: Proportion test results for sample size 75

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (85.00%) Combination 1, 3, 4 & 5 (17.00%)	M1.1	ML	0.00000	Significant
CI (85.00%) Combination 2 (70.50%)	M1.1	ML	0.00024	Significant
CI (76.00%) Combination 1, 3, 4 & 5 (9.50%)	M1.2	ML	0.00000	Significant
CI (76.00%) Combination 2 (48.50%)	M1.2	ML	0.00000	Significant
CI (67.00%) Combination 1, 4 & 5 (37.50%)	M1.1	GLS	0.00000	Significant
CI (67.00%) Combination 2 (39.50%)	M1.1	GLS	0.00000	Significant
CI (67.00%) Combination 3 (57.00%)	M1.1	GLS	0.01970	Significant
CI (78.50%) Combination 1, 4 & 5 (46.50%)	M1.2	GLS	0.00000	Significant
CI (78.50%) Combination 2 (66.50%)	M1.2	GLS	0.00357	Significant
CI (78.50%) Combination 3 (52.00%)	M1.2	GLS	0.00000	Significant

At sample size 75, the CI was the best performing index, correctly specifying 85%, 76%, 67%, and 78.5% of the misspecified models M1.1 and M1.2 for both estimation methods, ML and GLS, respectively. Combination 2 performed second-best for models M1.1 for ML, as well as M1.2 for ML and GLS estimation. However, for M1.1 under ML estimation, combination 3 is the second-best performing index. The same pattern is observed for sample size 100 (Figure 5.7(c)). The results of the proportion tests for sample size 75 can be found in Table 5.23. For M1.1, under ML estimation, the difference between the CI and all the combinations is found to be significant (Table 5.23).

For M1.2, under ML estimation, the performance difference between the CI and all the combinations are shown to be statistically significant. For M1.1, under GLS estimation, the performance difference between the CI and all the combinations are shown to be statistically significant. For M1.2, under GLS estimation, the difference in performance of the CI and all the combinations are assessed and shown to be statistically significant.

For models M1.1 and M1.2, under ML estimation at sample size 100, the CI correctly classified 92.5% and 90% of the models, respectively (Figure 5.7(c)). Combinations 1, 3, 4 and 5 correctly classified 12.5% and 10.5% of the models, respectively. This is attributed to the presence of SRMR within these combinations, as it was shown earlier in this section that the SRMR did not adequately detect model misspecification under ML estimation. The difference in performance between the CI and all the combinations, for M1.1 and M1.2, under ML estimation, were found to be statistically significant (Table 5.24).

Table 5.24: Proportion test results for sample size 100

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (92.50%) Combination 1, 3, 4 & 5 (12.50%)	M1.1	ML	0.00000	Significant
CI (92.50%) Combination 2 (75.00%)			0.00000	Significant
CI (90.00%) Combination 1, 3, 4 & 5 (10.50%)	M1.2		0.00000	Significant
CI (90.00%) Combination 2 (73.00%)			0.00000	Significant
CI (67.84%) Combination 1, 4 & 5 (40.20%)	M1.1	GLS	0.00000	Significant
CI (67.84%) Combination 2 (42.71%)			0.00000	Significant
CI (67.84%) Combination 3 (50.75%)			0.00026	Significant
CI (78.89%) Combination 1, 4 & 5 (38.19%)	M1.2		0.00000	Significant
CI (78.89%) Combination 2 (66.83%)			0.00336	Significant
CI (78.89%) Combination 3 (39.70%)			0.00000	Significant

For models M1.1, under GLS estimation at sample size 100, the CI performed best compared to the combination by correctly classifying 67.84% of the models. Combinations 1, 4 and 5 correctly classified 40.2% of the models, combination 2 classified 42.71% correctly and combination 3 classified 50.75% correctly. The difference in performance between the CI and the combinations are assessed and shown to be statistically significant (Table 5.24).

For models M1.2, under GLS estimation at sample size 100, the CI outperformed the combination by correctly classifying 78.89% of the models. Combinations 1, 4 and 5 correctly classified 38.19% of the models, combination 2 classified 66.83% correctly and combination 3 classified 39.70% correctly. Even though the RMSEA did not adequately detect misspecification under GLS estimation, the combination containing this value still performed second best in classifying the models. This is attributed to the presence of the TLI within the combination, which is shown to detect misspecification very well. The difference in performance between the CI and the combinations are assessed and shown to be statistically significant (Table 5.24).

As the sample size increases to 300, 500, and 1000, (Figures 5.7(d), (e), and (f)), under ML estimation, only combination 2 and the CI, correctly classified any of the models. This was since the SRMR failed to correctly specify any of the models under ML estimation at these sample sizes. This had an impact on the classifications of the combinations that contained SRMR. For sample size 300, the CI correctly classified 99% of the models, while combination 2 classified 92.33% of the models correctly. The CI significantly outperformed combination 2 for sample size 300 for model M1.1 (Table 5.25).

Table 5.25: Proportion test results for sample size 300

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (99.00%)	M1.1	ML	0.00000	Significant
Combination 1, 3, 4 & 5 (0.00%)				
CI (99.00%)			0.00000	Significant
Combination 2 (92.33%)				
CI (94.33%)	M1.2		0.00000	Significant
Combination 1, 3, 4 & 5 (0.00%)				
CI (94.33%)			0.00000	Significant
Combination 2 (60.00%)				
CI (47.49%)	M1.1	GLS	0.00000	Significant
Combination 1, 3, 4 & 5 (23.41%)				
CI (47.49%)			0.31207	Insignificant
Combination 2 (45.48%)				
CI (73.91%)	M1.2		0.00000	Significant
Combination 1, 3, 4 & 5 (28.43%)				
CI (73.91%)			0.42465	Insignificant
Combination 2 (73.24%)				

For model M1.2, the CI correctly classified 94.33% of the models and combination 2 correctly classified 60% of the models. The difference in performance is assessed and found to be statistically significant (Table 5.25).

Under GLS estimation, at sample sizes 300, 500, and 1000, all the combinations correctly classified the models to a certain extent. The CI outperformed the next best combination, combination 2 by 2.01% and 1.01% for misspecification model M1.1, for sample sizes 300 and 500. These differences in performances are assessed and found to be statistically insignificant (Table 5.25 and 5.26, respectively). The performance difference between the CI and the remaining combinations are found to be significant for M1.1 for sample sizes 300 and 500.

For model M1.2, under GLS estimation the CI performed best by correctly classifying 73.91% and 75.84% of the models for sample sizes 300 and 500, respectively. Combination 2 has the second highest performance rate at 73.24% and 75.5%, respectively. The remaining combinations correctly classified 28.43% and 30.20%, respectively. The difference in performances are assessed and the results captured in Tables 5.25 and 5.26, respectively. The difference between the CI and combination 2 are found to be insignificant at both sample sizes. The difference between the CI and the remaining combinations are found to be significant at sample size 300 and 500.

Table 5.26: Proportion test results for sample size 500

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (100.00%) Combination 1, 3, 4 & 5 (0.00%)	M1.1	ML	0.00000	Significant
CI (100.00%) Combination 2 (96.67%)	M1.1	ML	0.00071	Significant
CI (98.67%) Combination 1, 3, 4 & 5 (0.00%)	M1.2	ML	0.00000	Significant
CI (98.67%) Combination 2 (61.67%)	M1.2	ML	0.00000	Significant
CI (43.29%) Combination 1, 3, 4 & 5 (22.48%)	M1.1	GLS	0.00000	Significant
CI (43.29%) Combination 2 (42.28%)	M1.1	GLS	0.40129	Insignificant
CI (75.84%) Combination 1, 3, 4 & 5 (30.20%)	M1.2	GLS	0.00000	Significant
CI (75.84%) Combination 2 (75.50%)	M1.2	GLS	0.46017	Insignificant

For model M1.1, under ML estimation for sample size 500, the CI correctly classified 100% of the models and combination 2 correctly classified 96.67% of the models. This difference is assessed and found to be significant (Table 5.26). For model M1.2, under ML estimation for sample size 500, the CI correctly classified 98.67% of the models while combination 2 classified 61.67% correctly. The difference in performance is assessed and found to be significant (Table 5.26).

Table 5.27: Proportion test results for sample size 1000

Comparison (% correct)	Model	Estimation	P-value	Outcome
CI (100.00%) Combination 1, 3, 4 & 5 (0.00%)	M1.1	ML	0.00000	Significant
CI (100.00%) Combination 2 (99.33%)	M1.1	ML	0.07780	Insignificant
CI (100.00%) Combination 1, 3, 4 & 5 (0.00%)	M1.2	ML	0.00000	Significant
CI (100.00%) Combination 2 (69.007%)	M1.2	ML	0.00000	Significant
CI (37.33%) Combination 1, 3, 4 & 5 (23.33%)	M1.1	GLS	0.00010	Significant
CI (37.33%) Combination 2 (37.33%)	M1.1	GLS	0.50000	Insignificant
CI (69.33%) Combination 1, 3, 4 & 5 (32.33%)	M1.2	GLS	0.00000	Significant
CI (69.33%) Combination 2 (69.33%)	M1.2	GLS	0.50000	Insignificant

Table 5.27 contains the results of the proportion tests for model M1.1 under ML estimation at sample size 1000. The CI performs best by classifying 100% of the models correctly and combination 2 classified 99.33% of the models. The remaining combinations failed to correctly classify any of the models. Once again this is because SRMR failed to correctly classify any of the models and each of

these combinations that failed to correctly classify any of the models contained the SRMR. The difference in performance between the CI and combination 2 is tested and found to be insignificant.

Table 5.27 contains the results for model M1.2 under ML estimation for sample size 1000. For this model, the CI performs best by correctly classifying 100% of the models. The combination that performs second-best is combination 2 that correctly classifies 69% of the models. The remaining combinations fail to correctly classify any of the models as each of these combinations contain the SRMR. The difference in performance between the CI and combination 2 is tested and found to be significant.

At sample size 1000, under GLS estimation, the CI matched the performance of combination 2 for misspecification model M1.1. The CI outperformed the remaining combinations by 14%. The difference in performance between the CI and combinations 1, 3, 4 and 5 is tested and found to be significant (Table 5.27).

For model M1.2 under GLS estimation at sample size 1000, the CI again matched the performance of combination 2, both correctly classifying 69.33% of the models. The remaining combinations classified 32.33% of the models correctly. The performance difference between the CI and combinations 1, 3, 4 and 5 is tested and found to be statistically significant (Table 5.27).

Overall the CI performs better than the index combinations in detecting misspecified models under both estimation methods. The CI matched the performance of combination 2 for sample size 1000 under GLS estimation for both models however it outperformed the remaining combinations. The CI outperformed all the combinations for the remaining sample sizes and estimation methods. When these performance differences were tested for statistical significance, five of the comparisons are found to be insignificant. The insignificant differences are found in models M1.1 and M1.2 under GLS estimation at sample size 300 and 500 as well as for model M1.1 under ML estimation at sample size 1000.

The results of the misspecified model provide validation to why reporting multiple indices can be problematic. As seen in the larger sample sizes under ML estimation, the combinations that contained SRMR failed to correctly classify any of the models' due to SRMR failing to classify the models correctly. Therefore, if the traditional assessment method is employed the overall results can be misleading depending on the indices chosen. The difference in performance between the CI and the index combinations in these situations provides strong evidence to suggest that using a composite index such as the CI may result in more accurate results.

5.5 Summary

This chapter discussed the results of the simulation study that was undertaken to assess whether the CI proposed in this study performed better than the traditional assessment method. The traditional assessment method requires reporting multiple goodness-of-fit indices. In order for a model to be classified as a good fit, each index reported should indicate a good fit based on the associated cut-off criteria.

This simulation study was separated into three experimental conditions that were alternated during the assessment of the CI and index combinations performance. These were sample size, estimation method and model misspecification. For each condition, the performance of the CI and index combinations was calculated. The performance was measured as the number of times that CI or combination correctly classified the model. In the case of the index combinations, this would require all the indices within the combination to correctly classify the model. The discrepancy between the average index values and the combination performances are attributed to this requirement.

Overall the CI outperformed all the combinations included in this study. The only exception was in sample size 50, under GLS estimation, where combination 2 outperformed the CI. This was explained by the fact that the SRMR failed to correctly classify any of the models, thus having an impact on the calculation of the CI. Under ML estimation, at sample sizes 300, 500, and 1000, the CI matched the performance of all the combinations used in this study, scoring 100% performance. The CI also matched the performance of all the combinations when assessing misspecification models M1.1 and M1.2, under GLS estimation at sample size 1000. For the misspecified models M1.1 and M1.2 under ML estimation at sample sizes 300, 500 and 1000, combinations 1, 3, 4 and 5 all failed to correctly classify any of the models. This was due to the SRMR failing to classify any models correctly. The SRMR classified all the models as having a good model fit when, as the models were misspecified, the fit should have been classified as bad. The reason that this result did not affect the calculation of the CI in this case was due to results of the TLI, CFI and RMSEA. The results of these indices were low enough to compensate for the high results of the SRMR, thus the calculation of the CI was not significantly affected.

By reviewing the results of the simulation study there seems to be empirical evidence to suggest that using a composite index performs better when assessing model fit, in comparison to the traditional assessment method. This suggests that using a composite index reflects the mathematical property "*the sum is greater than the sum of its parts*". Further study into the performance of composite indices is required to validate these results.

Chapter 6 Case Study

6.1 Introduction

The results in chapter five provided guidance on goodness-of-fit index selection for SEM modelling. To test the CI in a practical setting, data from an actual case study was required. In collaboration with the school of Industrial Psychology and Human Resources at Nelson Mandela University, data were collected for a study on the factors pertaining to talent management strategies, within a South African context.

The case study results are obtained using an ordered and a randomised questionnaire design. The results of each questionnaire type are compared and discussed. This chapter will cover the literature relating to talent management strategies (section 6.2) and factor analysis (section 6.3). The literature on SEM analysis was covered in chapter two. The remaining sections cover the methodology used in the case study (section 6.4), the results of the case study (section 6.5), and finally the comparison of results from the ordered and randomised questionnaire designs (section 6.6). As this study is focussed on assessing model fit, the main results reported for the case study refers to the adequacy of model fit.

6.2 Talent management strategies

In recent years, due to global economic changes, companies have started to realise the need for talent management strategies. Regardless of this consensus, there is a lack of theoretical framework regarding talent management and talent management strategies (Al Ariss, Cascio, and Paauwe, 2014). Collings and Mellahi (2009) define strategic talent management as the methods by which companies identify strategic areas within the business that will result in obtaining a competitive advantage. This includes developing a group of highly talented individuals, with great potential, that perform well. These individuals should then be used to fill the positions within the strategic areas identified by the company (Collings and Mellahi, 2009). There are three mainstream interpretations of talent management (Collings and Mellahi, 2009). These are described as:

- A new term of common human resource practices.
- A reference to practices relating to succession planning.
- The term refers more to the general management of talented employees.

The interpretations of talent management have led to the development of talent management strategies (Al Ariss et al., 2014). Improving the use of talent management strategies leads to an increased number of quality employees with improved skills, which in turn leads to an increase in innovative ability and job satisfaction. This will in turn lead to an increase in talent retention rate (Collings and Mellahi, 2009).

The benefits of talent management strategies can only truly be recognised by integrating these strategies into the corporate strategies (Bethke-Langenegger, Mahler and Staffelbach, 2011).

If talented employees feel appreciated within the company framework this would lead to improved motivation and loyalty to the company, resulting in better work quality and thus, higher customer satisfaction (Collings and Mellahi, 2009). This will then have a positive influence on the company's profit levels.

To attract talented employees, companies need to ensure that incentives are provided in line with the needs of the employees. Highly talented employees tend to favour positions that are challenging and provide opportunities for further advancements within the company (Bethke-Langenegger et al., 2011). Further reading on talent management strategies can be found in the work by Al Ariss et al. (2014) and Collings and Mellahi (2009).

6.3 Factor analysis (FA)

Factor analysis is a multivariate statistical technique used to identify latent factors. The basis of factor analysis is that a set of underlying or latent factors $f_1,\ldots, f_k$ account for the intercorrelations between observed variables $x_1,\ldots,x_p$, where $p > k$. This leads to a situation in which should factors be removed from the observed variables, there is no longer correlation between the observed variables (Jöreskog, 1978). There are many situations in which factor analysis is useful. Examples of these situations include a researcher wanting to identify the factor structure of a dataset, or to reduce a dataset into a more manageable size (Comrey and Lee, 1992, Williams, Brown, and Onsman, 2010). The types of factor analysis are discussed in the following sections.

6.3.1 Types of factor analysis

There are two types of factor analysis that are used to identify latent factors, namely, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The first form of factor analysis, EFA, is a data driven approach used to identify an empirical set of latent factors. The second form of factor analysis, CFA, is used by researchers to test a theoretical model (Henson and Roberts, 2006). This is done by imposing a factor structure onto the data and assessing the model fit using goodness-of-fit indices. The CFA process forms the initial steps of the SEM process.

6.3.2 Theory of EFA

EFA is a technique that is used to identify the underlying structure among observed variables, often referred to as items, within the analysis (Hair et al., 2014). It analyses the correlation structure among many items and summarises the highly-correlated items into groups, namely latent factors (Hayton, Allen, and Scarpello, 2004). These factors represent the dimension of the data. Another use of EFA is for data reduction in which representative variables are identified from a larger set of variables or a new set of variables is created to replace the original set of variables (Hair et al., 2014).

When deciding on the appropriate form of factor analysis to use, EFA is used when there is no prior knowledge or theoretical model regarding the factor structure of the data. Conducting a CFA on each plausible model would be a lengthy process due to the number of alternative models possible (Fabrigar, Wegener, MacCullum, & Strahan, 1999). Therefore, EFA is a time efficient form of factor analysis and the more accurate form as it removes the likelihood of not considering all plausible models (Fabrigar et al., 1999). The latent variables identified within the EFA can then be used in further forms of analysis.

6.3.3 Process of EFA

The process of performing an EFA involves six main steps (Hair et al., 2014). These steps are briefly discussed below and summarised in Figure 6.1. For further information on EFA see Fabrigar et al. 1999; Williams et al. 2010; and, Hair et al. 2014.

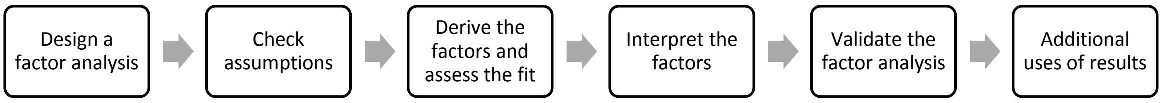


Figure 6.1: EFA process

Step 1: Design a factor analysis

The first stage in EFA involves the design of the analysis. An EFA should be designed by considering the questionnaire type used to collect data, the sample size required and the items used within the questionnaire. Another property to consider is the item to factor ratio, which should be at least 3:1 (preferably 4:1), that is possible relative to the number of items used (Fabrigar et al., 1999). Finally, the sample size is evaluated with regards to the absolute sample size and its relationship with the number of variables in the analysis (Fabrigar et al., 1999). Ideally, a minimum ratio of five respondents to each variable should be obtained.

Step 2: Check assumptions

The second stage of EFA consists of reviewing the underlying assumptions of the analysis. From a conceptual standpoint, the assumption that there exists some underlying structure in the selected variables must hold (Hair et al., 2014). From a statistical point, the overall measure of intercorrelation needs to be assessed, this is done using the following methods. The correlation matrix should be assessed to identify whether there are sufficient correlations above 0.3 (Williams et al., 2010). The

Bartlett's Test of Sphericity needs to be performed to test for the presence of correlations between the items, a significant result is desired. Finally, the Keiser-Meyer-Olkin measure of sampling adequacy (KMO MSA), which measures the proportion of variance of the items that may be caused by an underlying factor structure, needs to be assessed (Williams et al., 2010).

Step 3: Derive the factors and assess the fit

The third stage of EFA is the derivation of the factors. The two methods of factor extraction are common factor analysis and component analysis (Costello and Osborne, 2005). The selection of the extraction method is based on two considerations. Firstly, the objectives of the factor analysis are considered. Secondly, the amount of *priori* information available regarding the variance of the variables is considered (Costello and Osborne, 2005). The total variance of a variable is made up of the common, specific and error variance. The common factor method is used when only the common variance needs to be considered (Costello and Osborne, 2005). It is used in the identification of the dimensions of the factor structure. Component analysis (principal component analysis) considers the total variance (Costello and Osborne, 2005). It is mainly used when data reduction is the objective of the study.

There are certain criteria that can be used to identify the number of factors to be extracted (Williams et al., 2010; Skrondal and Rabe-Hesketh, 2004). The Keiser criterion can be used by extracting factors based on whether the respective eigenvalues are greater than one. The percentage of variance explained can also be used to determine the number of factors to extract. A cumulative percentage of between 50-60% is considered satisfactory for social sciences, while approximately 95% is required for the natural sciences. The scree plot is a graphical method that can be used to determine the factors extracted. Finally, the number of factors extracted can be based upon a *priori* criterion.

Step 4: Interpret the factors

The fourth stage of EFA concerns the interpretation of the factors. To select and interpret the final factor solution, three fundamental processes can be performed. Firstly, the unrotated factor matrix is estimated. The next process is to rotate the factor matrix to obtain a factor matrix that is simpler to interpret. This is done by redistributing the variance between the factors (Williams et al., 2010). The matrix is rotated using orthogonal or oblique rotation based on the assumptions regarding the factors (Comery and Lee, 1992). The final process required is to interpret the factors by assessing the factor loadings and the communalities. A factor loading of a minimum of 0.30 is deemed to be significant however higher loadings are often preferred (Williams et al., 2010). This is because the factor loadings are proportional to the quantity of variance accounted for by the factor. The factor loading value chosen should consider the sample size being analysed, guidelines can be found in Hair et al. (2014).

The communality of each variable represents the amount of variance accounted for in the factor solution of each item (Hair et al. 2014). For a variable to contribute significant information to the explanation of a factor, its communality should be greater than 0.50. However, communalities of 0.30 or higher are accepted as significant (Hair et al. 2014).

Once the factors have been evaluated, a respecification of the model may be needed. This is done by removing certain items, an alternative rotation method chosen, a different number of factors extracted or the method of extraction changed (Hair et al. 2014). Once the final factor solution is satisfactory, each factor is interpreted by reviewing the variables loaded onto each specific factor (Williams et al. 2010). The variable with the highest loading on a factor is the variable that contributes the most information to the factor. Once interpreted, a relevant label is assigned to the factors (Williams et al. 2010).

Step 5: Validate the factor analysis

The fifth stage of EFA refers to the validation of the factor analysis. This is done through the assessment of the degree of generalisability of the model. This is done through assessing validity using CFA. The stability of the factor structure can be assessed by evaluating sample size and the ratio of cases per variable. Finally, the presence of outliers need to be assessed (Hair et al., 2014).

Step 6: Additional uses of the results

The final stage of EFA deals with the additional uses of the EFA results (Williams et al., 2010). Once completed the results can be used in one of the following ways; the variable with the highest factor loading can be used to represent its respective factor, a smaller set of variables can be obtained using summated scales as a replacement dataset for large datasets. Factor scores can be calculated and used in further studies, and finally, it can be used in SEM analyses (Hair et al., 2014).

6.4 Methodology

This case study is performed to identify the underlying models pertaining to talent attraction. It is also used to compare the two types of questionnaires to assess whether the item order influenced the results. The following section will discuss the questionnaires used within the study, the details of the data, the software used, and the analysis methodology undertaken.

6.4.1 Questionnaire design

Two types of questionnaires are used within the case study, one following an ordered questionnaire design and the second following a randomised questionnaire design. Both questionnaires contained the same items pertaining to factors associated with talent attraction. The items contained in both questionnaires are from validated sources (Parry and Urwin, 2009) and scored on a Likert scale. The items within the ordered questionnaire are listed in such a way that the items are grouped in their relevant factors. The items within the randomised questionnaires are listed using a random number generator to randomly order the items.

6.4.2 Data

The two questionnaires are distributed to respondents working in the corporate environment. The completed questionnaires are collected and the data captured. The data is cleaned to ensure all codes are correct and to identify outliers. Once cleaned the ordered questionnaire dataset contained 361 respondents and the randomised questionnaire dataset contained 779 respondents.

6.4.3 Software

During the analysis of this study, different software were used based on the capabilities of the software. There are numerous softwares that can be used for EFA and SEM. The software available to perform EFA include Statistica, R and SPSS. The software that can be used to perform SEM includes R, AMOS, LISREL, Minitab and EQS. Due to availability and use within the literature, SPSS 24.0 is chosen to run the EFA analysis and AMOS 24.0 is chosen to run the SEM analysis.

6.4.4 Analysis

The questionnaires are analysed using EFA and SEM. The factor analysis chosen for this study is EFA as there is no preconceived theoretical structure underlying the data. The structures obtained through the EFA analysis are used in the SEM analysis.

EFA methodology

Both datasets are imported into the SPSS statistical program. The EFA is completed using the following criteria. The KMO MSA and Bartlett's Test of Sphericity are computed to check the suitability of the data. The EFA analysis is then completed using the maximum likelihood estimation extraction method as this method has the best support in the literature.

The criteria used to determine the number of factors extracted is the Kaiser's eigenvalue criteria. The cumulative percentage of variance extracted is also reviewed. The type of rotation used in the analysis is oblique rotation. The data pertains to human behaviour; thus, it is safe to assume that there will be correlation between variables. The method of oblique rotation used is PROMAX rotation as it is the most commonly used method in the literature.

The factor loadings are set to 0.40 to be deemed statistically significant. Once the factor solution is obtained, the communalities are reviewed to ensure that all communalities are above 0.3 to ensure that significant variation is explained by the model. Once the analysis is run the relative output is reviewed. This general methodology is followed to run the EFA analysis on both datasets.

Ordered questionnaire methodology

The initial solution is computed following the general methodology stated previously. The number of factors extracted is 5, based on the Kaiser's eigenvalue criteria. From the output, the pattern matrix is assessed and it is found that factors 3 to 5 only had 2 items loaded onto each factor. As stated in the literature, a minimum of 3, preferably 4, items should load onto a factor for it to be significant (Hair et al., 2014). Therefore, fewer factors need to be extracted.

A second analysis is completed setting the number of factors extracted to 4, *ceteris paribus*. The output of the second analysis is reviewed. The pattern matrix is reviewed and it revealed that factors 3 and 4 had only 3 items loaded onto each factor. While per the literature, the minimum ratio of items to factors that should be used is 3:1, a ratio of 4:1 is a more desirable ratio. Therefore, another analysis is performed with 3 factors, *ceteris paribus.* The output is reviewed and the pattern matrix obtained shows that the 3 extracted factors each had an item to factor ratio of 4:1 or greater. This indicates that an appropriate pattern matrix.

Each of the factors are then assessed for reliability by calculating the Cronbach's alpha values. For the factor to be reliable, the Cronbach's alpha value should be greater than 0.70. Once the factors are found to be reliable and have been interpreted, the pattern matrix is used in the SEM analysis which will be discussed in the following section.

Randomised questionnaire methodology

The initial EFA solution is computed following the general methodology stated in the EFA methodology section. The initial number of factors extracted is 3 with an adequate item to factor ratio. No further EFA analyses are required. Each factor is then assessed for reliability using the Cronbach alphas. All factors are found to be reliable per the criteria stated previously. As the factors are found to be reliable, they are interpreted and the pattern matrix used in subsequent SEM analyses.

SEM methodology

The pattern matrix obtained in the EFA analysis is used to draw the CFA model in AMOS. Once the CFA model is constructed, the analysis is run. The dataset is uploaded into AMOS and the analysis properties selected. Maximum likelihood estimation is chosen as it had the best support in the literature reviewed. The MI is set to 10 as only large MI are to be considered when reviewing the covariances (Byrne, 2010). The analysis is then run obtaining the standardised estimates. Once the analysis output is obtained, the model fit is assessed by reviewing the goodness-of-fit indices.

The absolute fit indices that are assessed include the χ^2 statistic, SRMR and RMSEA. These indices are assessed as they are the indices used in the development of the CI. The incremental fit indices used are the CFI and TLI indices. The CFI is chosen over the normed fit index (NFI) as it is less sensitive to model complexity. It is also one of the indices used in the CI.

The unstandardised path estimates are assessed for statistically significant estimates and standard residuals of less than |4.0|. The standardised path estimates are assessed as loadings of less than 0.5 should be reviewed as this suggests that the estimates are insignificant (Hair et al., 2014). If any respecifications are made, the new model is assessed following the same method as previously stated.

Once an adequate measurement model is obtained, the structural model is developed. This is done by incorporating a latent factor defined by the factors in the CFA model. The structural model is developed from the theoretical objective of the research.

The structural model developed is analysed using maximum likelihood estimation and the output is obtained. The structural model output is assessed using the same goodness-of-fit indices used to assess the CFA model. The path estimates of the model are assessed to identify insignificant estimates.

The MI values are assessed to obtain the significant covariances to be taken into consideration. Once modifications have been made, the new model is then analysed and the output assessed using the same measures stated above. Conclusions are then drawn regarding the model fit. The CI results are compared to the traditional methods to determine whether the CI indicates a different overall model fit result when compared to the index combinations. This comparison provides an indication of whether the CI is more successful at classifying model fit.

6.5 Results

Data is collected for both types of questionnaire design to assess whether a difference between the structure and model fit is observed depending on the questionnaire type used. A discussion of the differences observed follows the results of the case study.

6.5.1 Ordered questionnaire design

EFA analysis

The ordered questionnaire is administered and the data collected for analysis. The method defined in the section 6.4.4 is applied to the data. The results of the EFA are discussed next. From Table 6.1 the KMO MSA value of 0.906, a significant Bartlett's Test of Sphericity and a sample size of 361 indicates that EFA is appropriate to perform on the data.

Table 6.1: KMO MSA and Bartlett's test output

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.906
Bartlett's Test of Sphericity	Approx. Chi-Square	2955.543
	df	190
	Sig.	0.000

The EFA is completed and 3 factors are extracted (Table 6.3). From the pattern matrix (Table 6.2) items Q5, Q8, Q12, Q13 and Q19 are excluded due to insignificant factor loadings (appendix B). The total variance explained with 3 factors extracted is 43.298% (Table 6.3), a low but acceptable result in the social sciences.

Table 6.2: Pattern matrix with 3 factors

	Factor		
	1	2	3
Q1	0.590		
Q2	0.790		
Q3	0.845		
Q4	0.466		
Q5			
Q6	0.656		
Q7	0.430		
Q8			
Q9			0.955
Q10			0.498
Q11	0.476		
Q12			
Q13			
Q14			0.474
Q15			0.716
Q16		0.760	
Q17		0.978	
Q18		0.469	
Q19			
Q20		0.664	

As observed in Table 6.3, five factors have an eigen value of greater than 1.00. The EFA is run by extracting five factors but is not deemed adequate as two of the factors extracted did not satisfy the condition of having three or more items loaded on each factor (appendix B). Due to this condition violation, three factors are extracted.

Table 6.3: Total variance explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.215	36.074	36.074	6.641	33.207	33.207
2	1.515	7.577	43.651	1.079	5.395	38.602
3	1.479	7.393	51.045	0.939	4.697	43.298
4	1.251	6.255	57.3			
5	1.099	5.496	62.796			
6	0.884	4.421	67.217			

Each factor is assessed for reliability using Cronbach's alpha (Table 6.4). All factors are found to be reliable with an alpha value greater than 0.70.

Table 6.4: Reliability Statistics

Factor	**Cronbach alpha**
1	0.755
2	0.816
3	0.825

The factors obtained are interpreted by specialists in the field of Business and Human Resource Management, and labelled as Work Environment, Financial Security and Development, respectively. The pattern matrix can now be used for further SEM analysis.

SEM analysis

The pattern matrix developed above is now used to develop a CFA model. The initial CFA model is found in appendix C, along with its respective goodness-of-fit indices. The analysis is run based on the methodology described in the previous chapter. Upon review of the results of the initial measurement model, the model requires respecification. The covariance path between error 1 and 2 is included in the model as the MI was larger than 10.00, and upon review of the items, including this pathway is logical. The two items are "*Feeling personally valued and honoured by colleagues*" and "*Being encouraged to give my opinions and ideas*". It is logical that when employees feel valued by colleagues, these colleagues would consider their opinions or ideas within the work environment. The analysis is run on the modified CFA model (Figure 6.2).

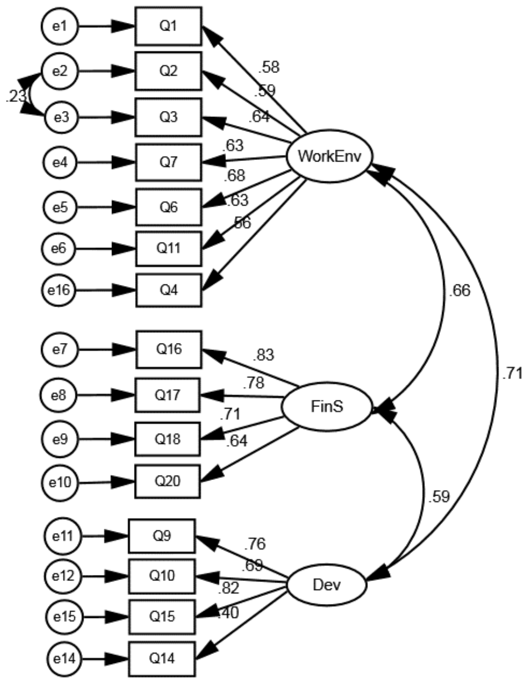


Figure 6.2: Modified measurement model

Table 6.5: Chi-square (χ^2) statistic for modified measurement model

Chi-square	283.179
Degrees of freedom	86
p-value	0.000

Table 6.6: Goodness-of-fit indices for modified measurement model

SRMR	0.065
RMSEA	0.080
CFI	0.903
TLI	0.882
CI	0.910

The chi-square (χ^2) test is significant, as seen in Table 6.5. The goodness-of-fit indices for the modified CFA model are in Table 6.6. The values of the CFI, TLI and CI are 0.903, 0.882 and 0.910, respectively. The values of the RMSEA and SRMR are 0.080 and 0.0654, respectively. The RMSEA value shows an adequate model fit as it is equal to the cut-off criteria that indicates an adequate fit. The SRMR indicates a good model fit as the value falls below the cut-off criteria indicating a good fit. The CFI and TLI, however indicated that the model fit was inadequate as both values fall below the associated cut-off criteria. The CI indicates a poor model fit as the value falls just below the cut-off of 0.9175 indicating an adequate fit. Therefore, using the traditional assessment method, this model is deemed a bad model fit as not all the indices indicated a good model fit. The CI indicated a poor model fit, albeit close to the adequate criteria.

The structural model is developed by loading the three factors, Work Environment, Financial Security and Development onto a single factor, Talent Attraction. The initial structural model can be found in appendix C. The analysis is run and the path estimates assessed. It is found that item Q14 had a standardised path estimate of 0.40, less than the required 0.50 (Hair et al., 2014). This item is removed from the model. The final structural model is observed in Figure 6.3. The analysis is run again and the output of the analysis found in Tables 6.7 and 6.8.

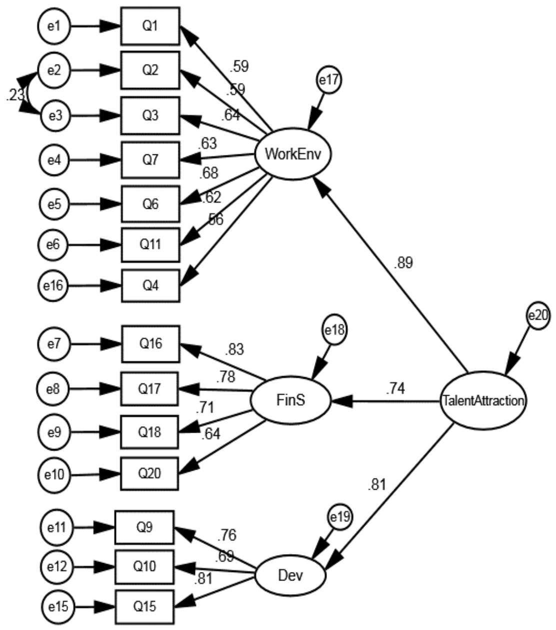


Figure 6.3: Final structural model

Table 6.7: Chi-square (χ^2) statistic for final structural model

Chi-square	240.513
Degrees of freedom	73
p-value	0.000

Table 6.8: Goodness-of-fit indices for final structural model

SRMR	0.064
RMSEA	0.080
CFI	0.915
TLI	0.894
CI	0.916

From Tables 6.7 and 6.8, the chi-square test is found to be significant and the model fit is assessed using the indices used to construct the CI. The CFI, TLI and CI values are 0.915, 0.894 and 0.916, respectively, while the RMSEA and SRMR values are 0.080 and 0.064, respectively. Once again, the RMSEA indicates an adequate model fit, while the SRMR indicates a good model fit. The CFI and TLI fall below the required cut-off values indicative of an adequate model fit, therefore indicating a poor model fit. The CI falls just slightly below the cut-off criteria, therefore indicating a poor model fit. Using the traditional assessment method, the overall model fit is deemed poor upon review of each index. The CI indicates a poor model fit although the value is so close to the cut-off value some researchers may deem it adequate.

Following the assessment of model fit, the path estimates are interpreted. The regression weights for the final structural model are found in Table 6.9. All the path estimates are significant and had a standard error of less than |4.0|. The standardized estimates (shown in Figure 6.3) can be interpreted as regression coefficients.

Table 6.9: Regression coefficients

			Estimate	S.E.	C.R.	P
WorkEnvironment	←	TalentAttraction	1			
FinancialSecurity	←	TalentAttraction	1.199	0.147	8.181	***
DevelopmentOpportunity	←	TalentAttraction	1.089	0.133	8.197	***

Summary

This case study was performed using the ordered questionnaire design data. Three factors were extracted during the EFA and these factors used within the SEM analysis. The SEM analysis was performed and the overall model fit using the traditional assessment method indicated a poor fit. The CI is 0.916, a result that falls below the cut-off value 0.9175. This indicates a poor model fit, although the value is so close to the adequate criteria that it may be interpreted as an adequate model fit.

6.5.2 Randomised questionnaire design

EFA analysis

The randomised questionnaire is administered and the data collected for analysis. The method previously discussed is applied to the data. The results of the EFA are now discussed. From Table 6.10 the KMO MSA value of 0.906, a significant Bartlett's Test of Sphericity and a sample size of 779 indicates that EFA is appropriate to perform on the data.

Table 6.10: KMO MSA and Bartlett's test output

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.906
Bartlett's Test of Sphericity	Approx. Chi-Square	4666.681
	df	105
	Sig.	0.000

The EFA is completed and 3 factors extracted using the Keiser criteria (Table 6.12). As seen in the pattern matrix (Table 6.11) item Q3 is excluded due to insignificant factor loading.

Table 6.11: Pattern matrix with 3 factors

	Factor		
	1	2	3
Q20		.606	
Q6			.401
Q9			.905
Q16		.876	
Q11	.546		
Q2	.643		
Q15			.628
Q10	.634		
Q18	.497		
Q14	.775		
Q17		.650	
Q7	.566		
Q4	.473		
Q1	.436		

The total variance explained with 3 factors extracted is 47.721%, a greater percentage compared to the result obtained from the ordered data (Table 6.12). This result is a low but acceptable in the social sciences. From Table 6.12, only three factors had an eigen value of greater than 1.00 in contrast to the five factors within the ordered questionnaire design study. Therefore, using the randomised design, the full factor structure indicated by the eigen values is used, however the items are different. These differences are discussed in section 6.5.3.

Table 6.12: Total Variance explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.230	41.532	41.532	5.679	37.858	37.858
2	1.289	8.591	50.123	.792	5.277	43.135
3	1.096	7.307	57.431	.688	4.586	47.721
4	.910	6.067	63.498			
5	.734	4.893	68.391			

Each factor extracted is assessed for reliability using Cronbach's alpha, the results of which can be seen in Table 6.13. Each factor showed adequate reliability per the Cronbach's alpha recommended values.

Table 6.13: Reliability Statistics

Factor	Cronbach alpha
1	0.828
2	0.736
3	0.745

The factors extracted are interpreted and labelled as Work Environment, Financial Security and Development, respectively. The pattern matrix is used in the SEM analysis.

SEM analysis

As with the ordered questionnaire analysis, the pattern matrix developed during the EFA is used to develop a CFA model. The analysis is run based on the methodology described in the previous section. A total of four covariance paths between errors are included in the model as the MI is larger than 10.00. The items associated with these errors are assessed to determine whether it is logical to free the pathway. The first pathway is between e1 and e2. The items associated with these errors are "*Teamwork*" and "*Feeling personally valued and honoured by colleagues*" which is deemed logical as if employees feel that being valued by colleagues is important, they are likely to feel that teamwork is important. The next set of errors are e2 and e4. The items associated with these errors are "*Feeling personally valued and honoured by colleagues*" and "*Recognition of my achievements*". This pathway is included as it is deemed logical that an employee that feels being valued is important would also find recognition important. The error e3 is associated with both errors e5 and e7. The item associated with e3 is "*Freedom and flexibility in my job*" and the item associated with e5 is "*Experiencing fun at work*". The item associated with e7 is "*Clear roles and responsibilities*". Both items associated with e5 and e7 can be logically associated with e3, as an employee that finds freedom and flexibility important may deem having clear roles and responsibilities, as well as having fun at work, important as it may lead to the employee feeling more secure within their job. The CFA model can be found in Figure 6.4 with the goodness-of-fit indices in Table 6.14 and 6.15.

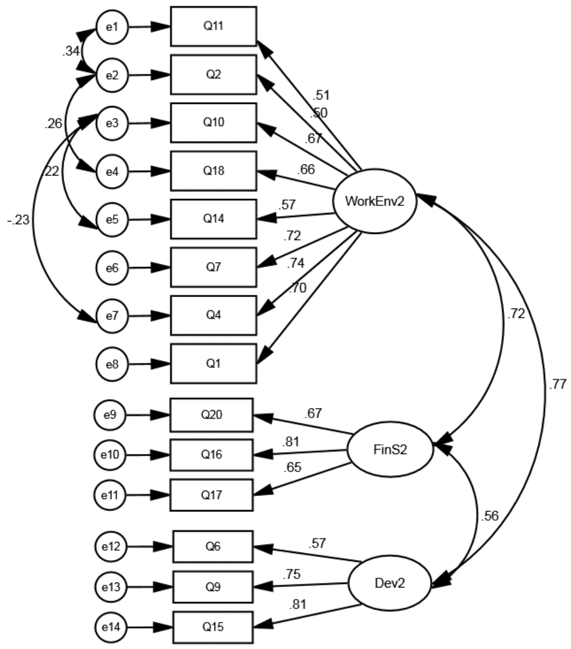


Figure 6.4: Measurement model for randomised design

Table 6.14: Chi-square (χ^2) statistic for measurement model

Chi-square	340.901
Degrees of freedom	70
p-value	0.000

Table 6.15: Goodness-of-fit indices for measurement model

SRMR	0.044
RMSEA	0.071
CFI	0.936
TLI	0.916
CI	0.934

The chi-square (χ^2) test was found to be significant. The CFI, TLI and CI values are 0.936, 0.916 and 0.934, respectively. The RMSEA and SRMR values are 0.071 and 0.0436, respectively. The CFI, CI and RMSEA indicate an adequate model fit. The SRMR indicates a good model fit while the TLI falls short of the cut-off indicating an adequate model fit, 0.95. Using the traditional assessment method, the overall model fit is deemed poor as only the SRMR indicates a good model fit and the CFI and RMSEA indicate an adequate fit. The low TLI result affects the overall model fit interpretation. The CI value indicates an overall adequate model fit. By comparing to the model to the validated scales, this model should reflect an adequate fit.

The structural model is developed by loading the three factors onto one factor, namely Talent Attraction. The analysis is run and the output obtained. The initial structural model can be found in appendix D along with the goodness-of-fit indices. Item Q2 is found to have a standardised path estimate of 0.495 which is lower than the required 0.50 minimum (Hair et al., 2014). The item is removed. The improved model is analysed and the output is discussed. The final structural model is observed in Figure 6.5. The chi-square test statistic and goodness-of-fit indices are in Tables 6.16 and 6.17.

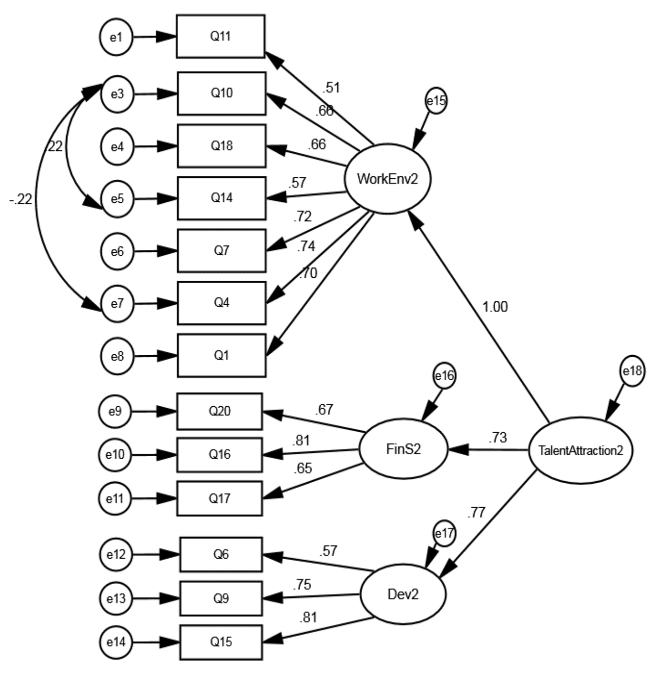


Figure 6.5: Final structural model for randomised design

Table 6.16: Chi-square (χ^2) statistic for final structural model for randomised design

Chi-square	291.820
Degrees of freedom	60
p-value	0.000

Table 6.17: Goodness-of-fit indices for final structural model for randomised model

SRMR	0.042
RMSEA	0.070
CFI	0.939
TLI	0.921
CI	0.937

The CFI, TLI and CI values are 0.936, 0.916 and 0.937, respectively, while the RMSEA and SRMR values were 0.071 and 0.044, respectively. The CFI, CI and RMSEA indicates an adequate model fit when comparing the values to the relative cut-off points. The SRMR indicates a good model fit relative to the associated cut-off point. The TLI, however, did not indicate an adequate model fit. The overall model fit, using the traditional assessment method, is therefore deemed poor. Using the CI to assess fit, the overall fit is deemed adequate. This result emphasises the benefit of using the CI to assess model fit, as the adequate fit result is in line with the theory.

As the model fit was significant, the model can be interpreted per the regression coefficients. Table 6.18 shows that all the factor path estimates are significant and the standard errors are less than |4.0|.

Table 6.18: Regression coefficients for randomised model

			Estimate	S.E.	C.R.	P
WorkEnvironment	←	TalentAttraction	1			
FinancialSecurity	←	TalentAttraction	0.699	0.062	11.207	***
DevelopmentOpportunity	←	TalentAttraction	0.935	0.073	12.854	***

The significant path estimates and appropriate standard errors indicate that conclusions can be drawn from the associated SEM model.

Summary

This case study was performed using the randomised questionnaire design data. Using the traditional assessment method, the overall model fit was deemed poor while the CI indicated an adequate model fit. By reviewing the theory, the adequate fit is more logical than a poor fit.

6.5.3 Comparison of ordered versus randomised design

As both questionnaires contained the same items, the way in which the items load onto the relevant factors could be compared. When compared, certain items in the randomised questionnaire failed to load on the original factors in the ordered questionnaire. It is found that there is a 50% item match for factor 1 (Work Environment), a 75% item match for factor 2 (Financial Security) and a 33% item match for factor 3 (Development).

Three items, Q10, Q18 and Q14 loaded onto different factors and one item, Q3, failed to load onto any factor in the randomised design study. These results are summarised in Table 6.19. Table 6.19 gives the item description, which factor it loaded on for each type of questionnaire and whether the change was logical or not. The changes are all deemed logical upon reviewing the definition of the factors.

Table 6.19: Item differences for ordered versus randomised design studies

Item Description	Ordered Questionnaire	Randomised Questionnaire	Logical/Illogical
Q3: "Being encouraged to give my opinions"	Factor 1 – Work Environment	Did not load	NA
Q10: "Freedom and flexibility in my job"	Factor 3 – Development	Factor 1 – Work Environment	Logical
Q18: "Recognition of my achievements"	Factor 2 – Financial Security	Factor 1 – Work Environment	Logical
Q14: "Experiencing fun at work"	Factor 3 – Development	Factor 1 – Work Environment	Logical

The item that did not load in the randomised questionnaire study is investigated. It is found that in the randomised questionnaire this item is the first item in the list and is followed by the item "*Job security*". However, in the ordered questionnaire this item is the third item in the list and had a moderate factor loading of 0.562. In the ordered questionnaire, this item is preceded by the item "*Feeling personally valued and honoured by colleagues*" and followed by the item "*Clear roles and responsibility*". By reviewing the order in which this item is listed, in the ordered questionnaire it is surrounded by "like" items which may have led respondents to answer Q3 in a similar method to the items around it. This is not the case in the randomised questionnaire. The item following Q3 is not a similar item and it subsequently did not load significantly. It can be argued that this indicates that the randomised questionnaire is a better methodology as it reduces the possibility of leading questions regarding the question order.

When comparing the overall model fit for the final structural models for each type of questionnaire, the randomised questionnaire produced a better overall model fit using the CI as the assessment method. The CI value for the randomised questionnaire design is 0.937 and 0.916 for the ordered questionnaire design. According to the CI value, the randomised questionnaire design produced the better model, this validates the findings of Malhotra (2006). The traditional assessment method deemed all the models inadequate.

The better overall model fit along with the more logical item loadings indicate that, in this case, the randomised questionnaire design produced a better model. To further investigate whether a randomised or ordered questionnaire produces better results, the two structures identified are imposed on the opposing data and the results documented in the following two sections.

6.5.4 Ordered structure on randomised data

The model obtained from the ordered questionnaire design is imposed on the randomised data to assess whether the model fit is consistent with new data and whether randomised data provides better results. If the imposed model is "true", the results produced should indicate a good model fit for all the data used to some extent. The results are discussed. The ordered structure (Figure 6.6) is imposed on the randomised data and the model fit can be seen in Tables 6.20 and 6.21.

The CFI, TLI and CI values are 0.849, 0.811 and 0.872, respectively. The RMSEA and SRMR values are 0.106 and 0.066, respectively. According to the SRMR result, the model represents a good model fit, with value below the predefined cut-off values. The results of the RMSEA, CFI and TLI indicate that the model has a poor model fit. Therefore, using the traditional assessment method the overall model fit indicates a bad fit. The CI value indicates a poor model fit as well.

The path estimates are all found to be significant with standard errors within the acceptable range. The standardised path estimates showed a positive relationship between factors. The regression weights can be found in appendix E.

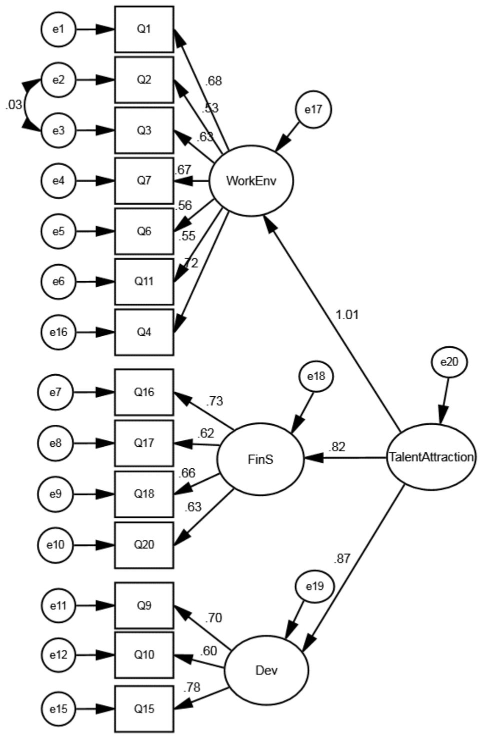


Figure 6.6: Ordered structural model on randomised data

Table 6.20: Chi-square (χ^2) statistic for ordered structure on randomised data

Chi-square	709.893
Degrees of freedom	73
p-value	0.000

Table 6.21: Goodness-of-fit indices for ordered structure on randomised data

SRMR	0.066
RMSEA	0.106
CFI	0.849
TLI	0.811
CI	0.872

6.5.5 Randomised structure on ordered data

The model obtained from the randomised questionnaire design is imposed on the ordered data. This model was deemed the more accurate model and therefore it is imposed on the ordered data to determine whether the results are consistent for this data. The results can be found in Tables 6.22 and 6.23.

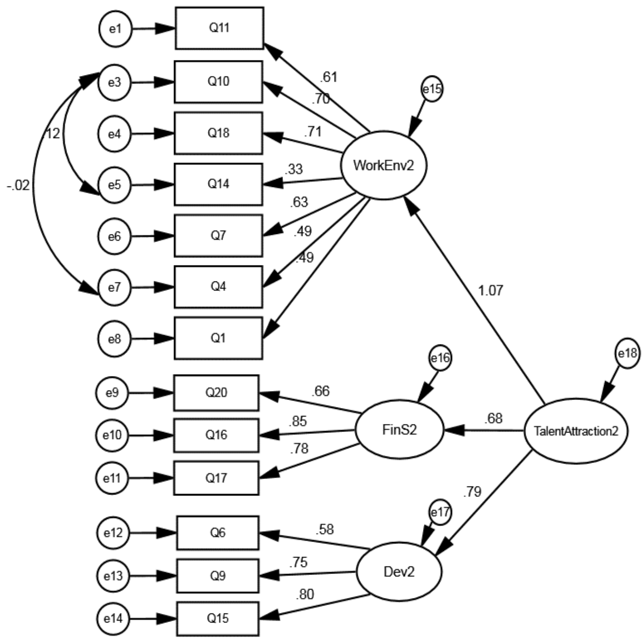


Figure 6.7: Randomised structure on ordered data

Table 6.22: Chi-square (χ^2) statistic for randomised structure on ordered data

Chi-square	262.381
Degrees of freedom	60
p-value	0.000

Table 6.23: Goodness-of-fit indices for randomised structure on ordered data

SRMR	0.066
RMSEA	0.097
CFI	0.881
TLI	0.846
CI	0.891

The CFI, TLI and RMSEA values indicate that the model fit is poor when compared to the respective cut-off criteria. The SRMR value indicates that the model has a good fit. Therefore, using the traditional assessment method the overall model fit indicates a poor fit. The CI value indicates a poor model fit. For this model, the traditional assessment method and CI reach the same conclusion.

The path estimates are all found to be significant with standard errors within the acceptable range. The standardised path estimates showed a positive relationship between factors as seen in Figure 6.7. The regression weights can be found in appendix F.

6.6 Summary

The case study was completed using both an ordered and randomised design questionnaires. Each model was analysed and the randomised design study appeared to produce a more appropriate model when reviewing the fit indices and the EFA structure. The EFA structure produced followed a more logical pattern when reviewing the items within each factor. This indicates that using a randomised design questionnaire appeared to produce more logical results, however, further testing into this concept is required.

The structures obtained for each case study was then imposed on the opposing dataset and the results reviewed. This was done to assess whether the results observed for the ordered and randomised models were consistent with different data. As the randomised model was deemed the more accurate and logical model, the model fit should be consistent using both sets of data. Using the CI values as a method of comparison for each respective model, the randomised structure that was imposed on the ordered data showed an inconsistent result when compared to the randomised questionnaire. This may imply that data collected using an ordered questionnaire design differs from data collected using a randomised questionnaire. Further work needs to be done to confirm this hypothesis. The ordered questionnaire model imposed on the randomised data was also consistent as it indicated a poor model fit, the same overall classification as seen in the ordered model.

Chapter 7 Conclusions and Recommendation

The primary aim of this study was to construct a composite goodness-of-fit index (CI) that would provide a better method of assessing model fit in SEM analyses. The CI was constructed using four goodness-of-fit indices of which two were absolute fit indices and two were incremental fit indices. The absolute fit indices used were the RMSEA and SRMR, which were reversed to maintain the directionality of the CI. The two incremental fit indices included were the TLI and CFI. The CI was linearly aggregated using equal weighting.

The second objective of the study was to assess the performance of the CI when compared to the traditional method of assessing model fit. The traditional method of assessing model fit is to report three to four goodness-of-fit indices, and each goodness-of-fit index reported should show a good model fit using the respective cut off values. Therefore, a total of five combinations were compared to the CI. These were made up of all the possible combinations, that could be formed using the indices CFI, TLI, RMSEA, and SRMR, while ensuring a minimum of three indices were reported as suggested in the literature.

The performance of the CI and each combination was assessed by identifying the number of models that were classified correctly using a Monte Carlo simulation study. The study also investigated the performance of the CI under varying experimental conditions. The conditions investigated were sample size, estimation methods and model misspecification. These percentages were compared to identify whether the CI outperformed the combinations and tested for statistical significance.

Sample sizes 50, 75, 100, 300, 500, and 1000 were investigated in this study to identify whether the use of a CI would show an improved performance compared to the different combinations. Sample size was investigated as it was one of the factors that had a great effect on assessing model fit in the literature. The CI showed a performance of 50% in the classification of the models with sample size 50, which was greater than the performance of the other combinations. Under sample size 75, the CI correctly classified 75.5% of the models, which was a better performance than the other combinations. For sample size 100, the CI resulted in a performance of 92.5%, which was greater than the classification percentages indicated by the traditional combinations. For sample sizes 300, 500, and 1000 the CI matched the performance of the traditional combinations, at a 100% performance. Therefore, the CI showed an improved performance at small sample sizes while maintaining a perfect performance at large sample sizes.

The estimation methods investigated in this study were ML and GLS. The results for the ML estimation method were the same as the sample size results above due to ML being the default estimation method in SEM analyses. Under GLS estimation, at sample size 50, the CI did not outperform combination 2 due to the SRMR failing to correctly classify any models. Under sample size 75, the CI (68%) outperformed all the combinations however the difference was insignificant with combination 2. The same pattern was observed under sample size 100. The combination 2 performed well at small sample sizes under GLS estimation. This combination consisted of the CFI, TLI, and RMSEA. This suggests that the SRMR does not perform adequately at small sample sizes under GLS estimation. For large sample sizes, 300, 500, and 1000, under GLS estimation, the CI significantly outperformed all the traditional combinations. Overall, the CI either equalled or outperformed the traditional indices at every sample size under ML estimation. Under GLS estimation, the CI outperformed the traditional combinations for all but one combination. Only combination 2, at sample size 50, outperformed the CI. However, two of the comparisons indicating that the CI outperformed combination 2 were found to be insignificant. This suggests that the CI performs better than the traditional combinations for both estimation methods investigated.

Two misspecified models were investigated at each sample size and estimation method. The misspecified model results for ML estimation are summarised in Table 7.1 according to whether the CI significantly outperformed the index combinations or not. The misspecified model results for GLS estimation are summarised in Table 7.2 using the same criteria as in Table 7.1.

Table 7.1: Summary of misspecified results for ML estimation

	M1.1					**M1.2**				
Sample size	**Combinations of indices**									
	1	2	3	4	5	1	2	3	4	5
50	Significant					Significant				
75	Significant					Significant				
100	Significant					Significant				
300	Significant					Significant				
500	Significant					Significant				
1000	Sig.	Insig.	Significant			Significant				

From Table 7.1, it is observed that the CI significantly outperformed all the combinations for both misspecified models under ML estimation, for sample sizes 50 to 500. For sample size 1000, the performance difference between the CI and combination 2 is observed to be insignificant for model M1.1 under ML estimation. The remaining results showed a significant difference between the CI and index combinations for both misspecified models.

Table 7.2: Summary of misspecified results for GLS estimation

	M1.1					M1.2				
Sample size	**Combinations of indices**									
	1	2	3	4	5	1	2	3	4	5
50	Significant		Insig.	Significant		Significant				
75	Significant					Significant				
100	Significant					Significant				
300	Sig.	Insig.	Significant			Sig.	Insig.	Significant		
500	Sig.	Insig.	Significant			Sig.	Insig.	Significant		
1000	Sig.	Insig.	Significant			Sig.	Insig.	Significant		

From Table 7.2 it is observed that for misspecified model M1.1 and M1.2, under GLS estimation, the difference between the CI and combination 2 is insignificant for sample sizes 300, 500 and 1000. The performance of the CI matched the performance of combination 2 at sample size 1000 for both misspecified models. At sample size 50, the difference between the CI and combination 3 is insignificant for misspecified model M1.1. For the remaining combinations at the differing sample sizes for models M1.1 and M1.2 are shown to be significant.

Overall, the CI either equalled or significantly outperformed all the combinations for the misspecified models under ML estimation. Under GLS estimation, the CI either equalled or significantly outperformed the traditional combinations for all but five combinations in the misspecified model comparisons. This suggests that the CI performs better than the traditional combinations when model misspecification is investigated. The results of this study suggest that the use of a CI may produce better results by correctly classifying more models, under experimental conditions such as sample size, estimation method and model misspecification. This supports the rationale of this study.

The final objective of the study was to use the CI in a case study to assess whether randomised and ordered questionnaire designs produce different factor structures. This objective was assessed in chapter 6 and it was found that randomised questionnaire designs produce a more logical factor structure. The CI was also applied to the case study analysis and the randomised design model resulted in a higher CI value which supported the result that this design produces more logical results.

While the results of this study suggest that using a CI may be a better method of assessing model fit, it is necessary to do further investigations into this topic. A recommendation would be to assess all logical combinations of the traditional goodness-of-fit indices to identify the optimal CI construction. Further work also needs to be done to investigate the performance of the CI with highly complex models and to investigate unequal weightings and nonlinear aggregation methods with a greater number of replications.

References

AL ARISS, A. A., CASCIO, W. F. & PAAUWE, J. 2014. Talent management: current theories and future research directions. *Journal of World Business,* 49**,** 173-179.

BAGOZZI, R. P. & YI, Y. 1988. On the evaluation of structural equation models. *Journal of the Academy of Marketing Science,* 16**,** 76-81.

BARR, G. D. I. & KANTOR, B. S. 2004. A criterion for comparing and selecting batsmen in limited overs cricket. *The Journal of the Operational Research Society,* 55**,** 1266-1274.

BEARDEN, W. O., SHARMA, S. & TEEL, J. E. 1982. Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of marketing research***,** 425-30.

BENTLER, P. M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin,* 107**,** 238-246.

BENTLER, P. M. & BONETT, D. G. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin,* 88**,** 588.

BETHKE-LANGENEGGER, P., MAHLER, P. & STAFFELBACH, B. 2011. Effectiveness of talent management strategies. . *European J. International Management,* 5**,** 524-539.

BOLLEN, K. A. 1989. A new incremental fit index for general structural equation models. *Sociological Methods & Research,* 17**,** 303-316.

BOLLEN, K. A. & NOBLE, M. D. 2011. Structural equation models and the quantification of behavior. *PNAS,* 108**,** 15639-15646.

BOLLEN, K. A. & STINE, R. A. 1992. Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research,* 21**,** 205-229.

BOOMSMA, A. 1982. The robustness of LISREL against small sample sizes in factor analysis models. *In:* JÖRESKOG, K. G. & WOLD, H. (eds.) *Systems Under Indirect Observation: Causality, Structure, Prediction (Part 1).* Amsterdam: North-Holland.

BOOMSMA, A. 1983. *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* PhD, University of Groningen.

BOOMSMA, A., HOYLE, R. H. & PANTER, A. T. 2012. The structural equation modeling research report. *Handbook of structural equation modeling***,** 341-358.

BROWNE, M. W. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology,* 37**,** 62-83.

BYRNE, B. M. 2010. *Structural equation modelling with AMOS: basic concepts, applications and programming.,* New York, Routledge.

CLIFF, N. 1983. Some cautions concerning the application of causal modeling methods. *Multivariate Behavioural Research,* 18**,** 115-126.

COLLINGS, D. G. & MELLAHI, K. 2009. Strategic talent management: a review and research agenda. *Human Resource Management Review,* 19**,** 304-313.

COSTELLO, A. B. & OSBORNE, J. 2005. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation,* 10**,** 1-9.

COUGHLAN, J., HOOPER, D. & MULLEN, M. 2008. Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods,* 6**,** 53-60.

CURRAN, P. J., BOLLEN, K. A., CHEN, F., PAXTON, P. & KIRBY, J. B. 2003. Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research,* 32**,** 208-252.

FABRIGAR, L. R., WEGENER, D. T., MACCULLUM, R. C. & STRAHAN, E. J. 1999. Evaluating the use of exploratory factor analysis in psychological research. *American Psychological Association Inc,* 4**,** 272-299.

FAN, X., THOMPSON, B. & WANG, L. 2009. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal,* 6**,** 56-83.

GERBING, D. W. & ANDERSON, J. C. 1992. Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research,* 21**,** 132-160.

GONZALO, J. & PITARAKIS, J. Y. 1998. Specification via model selection in vector error correction models. *Economic Letters,* 60**,** 321-328.

HAIR, J. F., BLACK, W. C., BABIN, B. J. & ANDERSON, R. E. 2014. *Multivariate data analysis.,* Essex, Pearson Education Limited.

HALLGREN, K. A. 2014. Conducting simulation studies in the R programming environment. *Tutor Quant Methods Psychol.,* 9**,** 43-60.

HAYTON, J. C., ALLEN, D. G. & SCARPELLO, V. 2004. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organisational Research Methods,* 7**,** 191-205.

HENSON, R. K. & ROBERTS, J. K. 2006. Use of exploratory factor analysis in published research. *Educational and Psychological Measurement,* 66**,** 393-416.

HINKIN, T. R. 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods,* 2**,** 104-121.

HOX, J. J. & BECHGER, T. M. 2007. An introduction to structural equation modeling. *Family Science Review,* 11**,** 354-373.

HU, L. & BENTLER, P. M. 1998. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods,* 3**,** 424-453.

HU, L. & BENTLER, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling,* 6**,** 1-55.

IACOBUCCI, D. 2009. Structural equations modeling: fit indices, sample size, and advanced topics. *Journal of Consumer Psychology,* 20**,** 90-98.

JAMES, L. R., MULAIK, S. A. & BRETT, J. M. 1982. Causal analysis: assumptions, models, and data. *SAGE Publications, Incorporated*.

JÖRESKOG, K. G. 1978. Structural Analysis of Covariance and Correlation Matrices. *Psychometrika,* 43**,** 443-477.

JÖRESKOG, K. G. & SÖRBOM, D. 1989. LISREL 7: A guide to the program and applications. *In:* SPSS (ed.).

KENNY, D. A. 2015. *Measuring model fit* [Online]. Available: http://davidakenny.net/cm/fit.htm [Accessed 27 October 2017].

KLINE, R. B. 2016. *Principles and practice of structural equation modeling,* New York, The Guilford Press.

LEE, S. 2015. Implementing a simulation study using multiple software packages for structural equation modeling. *SAGE Open***,** 1-16.

LEI, P. W. & WU, Q. 2012. Estimation in structural equation modeling. *In:* HOYLE, R. H. (ed.) *Handbook of structural equation modeling.* New York: Guilford Press.

MACCALLUM, R. C. & AUSTIN, J. T. 2000. Applications of structural equation modeling in psychological research. *Annual review of psychology,* 51**,** 201-226.

MALHOTRA, N. K. 2006. Questionnaire design and scale development. *The handbook of marketing research: Uses, misuses, and future advances.*

MARKLAND, D. 2006. The golden rule is that there are no golden rules: a commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences,* 42**,** 851-858.

MCDONALD, R. P. & HO, M. R. 2002. Principles and practice in reporting structural equation analyses. *Psychological Methods,* 7**,** 64-82.

MCDONALD, R. P. & MARSH, H. W. 1990. Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin,* 107**,** 247.

MULAIK, S. A., JAMES, L. R., VAN ALSTINE, J., BENNETT, N., LIND, S. & STILWELL, C. D. 1989. Evaluation of goodness-of-fit indices for structural equation models. *Quantitative Methods in Psychology,* 105**,** 430-445.

NACHTIGALL, C., KROEHNE, U., FUNKE, F. & STEYER, R. 2003. (Why) should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research Online,* 8**,** 1-22.

NARAYANAN, A. 2012. A review of eight software packages for structural equation modeling. *The American Statistician,* 66**,** 129-138.

NEALE, M. C., HUNTER, M. D., PRITIKIN, J. N., ZAHERY, M., BRICK, T. R., KIRKPATRICK, R. M., ESTABROOK, R., BATES, T. C., MAES, H. H. & BOKER, S. M. 2016. OpenMx 2.0: extended structural equation and statistical modeling.: Psychometrika.

PARRY, E. & URWIN, P. 2009. *Tapping into talent: The age factor and generation issues.,* London, Chartered Institute of Personnel and Development.

PAXTON, P., CURRAN, P. J., BOLLEN, K. A., KIRBY, J. & CHEN, F. 2001. Monte carlo experiments: design and implementation. *Structural Equation Modeling* 8**,** 287-312.

PORNPRASERTMANIT, S., MILLER, P., SCHOEMANN, A., QUICK, C. & JORGENSEN, T. 2016. Simulated structural equation modeling. 0.5-13 ed.

RATTRAY, J. & JONES, M. C. 2005. Essential elements of questionnaire design and development. *Journal of Clinical Nursing,* 16**,** 234-243.

SCHERMELLEH-ENGEL, K., MOOSBRUGGER, H. & MULLER, H. 2003. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online,* 8**,** 23-74.

SCHREIBER, J. B., NORA, A., STAGE, F. K., BARLOW, E. A. & KING, J. 2006. Reporting structural equation modelling and confirmatory factor anaysis results: A review. *The Journal of Educational Research,* 99**,** 323-337.

SCHUMACKER, R. E. & LOMAX, R. G. 2004. *A beginner's guide to structural equation modelling.,* New Jersey, Lawrence Erlbaum Associates.

SHAH, R. & GOLDSTEIN, S. M. 2006. Use of structural equation modeling in operations management research: looking back and forward. *Journal of Operations Management,* 24**,** 148-169.

SHEVLIN, M. & MILES, J. N. V. 1998. Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences,* 25**,** 85-90.

SKRONDAL, A. & RABE-HESKETH, S. 2004. *Generalized latent variable modelling: multilevel, longitudinal, and structural equation models,* Florida, Chapman & Hall/CRC.

STANGROOM, J. 2017. *Social science statistics* [Online]. Available: http://www.socscistatistics.com/tests/ztest/ [Accessed 2017].

SUHR, D. 2006. *Exploratory or confirmatory factor analysis?,* Cary: SAS Institute.

THOMPSON, B. 2000. Ten commandments of structural equation modeling. *In:* GRIMM, L. G. & YARNOLD, P. R. (eds.) *Reading and understanding more multivariate statistics.* Washington, DC: American Psychological Association.

WESTON, R. & GORE, P. A. 2006. A brief guide to structural equation modeling. *The counseling psychologist,* 34**,** 719-751.

WILLIAMS, B., BROWN, T. & ONSMAN, A. 2010. Exploratory factor analysis: a five-step guide for novices. *Australian Journal of Paramedicine,* 8**,** 1-13.

Appendix A

The following table gives an example of the parameter estimates generated for each replication of the simulation model at sample size 50. The important values are highlighted in red. Within the parameter estimates, the significance and standard errors of the regression weights should be reviewed.

Table A.1: Parameter estimate output example for sample size 50

Latent Variables:

	Estimate	**Standard Error**	**Z-value**	**P-value**
Factor 1 – x1	1.000			
Factor 1 – x2	0.501	0.188	2.662	0.008
Factor 1 – x3	0.730	0.191	3.824	0.000
Factor 1 – x4	0.991	0.222	4.469	0.000
Factor 1 – x5	0.717	0.180	3.992	0.000
Factor 1 – x6	1.697	0.282	6.009	0.000
Factor 1 – x7	0.931	0.196	4.757	0.000
Factor 2 – x8	1.000			
Factor 2 – x9	0.897	0.214	4.199	0.000
Factor 2 – x10	1.392	0.266	5.234	0.000
Factor 2 – x11	1.313	0.281	4.663	0.000
Factor 3 – x12	1.000			
Factor 3 – x13	0.720	0.214	3.364	0.001
Factor 3 – x14	0.892	0.265	3.371	0.001

Regressions:

	Estimate	**Standard Error**	**Z-value**	**P-value**
Factor 1 – Factor 2	0.344	0.156	2.203	0.028
Factor 1 – Factor 3	0.395	0.187	2.107	0.035
Factor 2 – Factor 3	0.475	0.210	2.261	0.024

Variances:

x1	0.319	0.074	4.284	0.000
x2	0.506	0.103	4.892	0.000
x3	0.427	0.090	4.725	0.000
x4	0.485	0.107	4.548	0.000
x5	0.364	0.078	4.687	0.000
x6	0.273	0.103	2.662	0.008
x7	0.343	0.077	4.430	0.000
x8	0.601	0.131	4.608	0.000
x9	0.458	0.100	4.582	0.000
x10	0.098	0.080	1.219	0.223
x11	0.594	0.141	4.206	0.000
x12	0.367	0.122	3.016	0.003
x13	0.315	0.083	3.799	0.000
x14	0.478	0.126	3.783	0.000
Factor 1	0.204	0.077	2.629	0.009
Factor 2	0.360	0.148	2.428	0.015
Factor 3	0.405	0.167	2.427	0.015

Appendix B

The following appendix provides the variance explained and pattern matrices for the EFA analysis on the ordered data.

Table B.1: Total variance explained for five factors extracted

Total Variance Explained						
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.215	36.074	36.074	6.670	33.350	33.350
2	1.515	7.577	43.651	1.090	5.450	38.799
3	1.479	7.393	51.045	.985	4.927	43.727
4	1.251	6.255	57.300	.677	3.383	47.109
5	1.099	5.496	62.796	.836	4.182	51.291
6	.884	4.421	67.217			
7	.706	3.530	70.746			
8	.679	3.393	74.140			
9	.628	3.138	77.277			
10	.589	2.944	80.222			
11	.584	2.919	83.140			
12	.513	2.564	85.704			
13	.457	2.286	87.990			
14	.442	2.212	90.202			
15	.420	2.101	92.302			
16	.379	1.895	94.198			
17	.357	1.783	95.981			
18	.288	1.442	97.422			
19	.273	1.365	98.787			
20	.243	1.213	100.000			

Table B.2: Five factor pattern matrix

	Factor				
	1	2	3	4	5
Q1	.543				
Q2	.590				
Q3	.646				
Q4	.690				
Q5	.524				
Q6	.639				
Q7					
Q8					
Q9			.878		
Q10					
Q11					
Q12					.694
Q13				.771	
Q14					.643
Q15			.629		
Q16		.754			
Q17		.867			
Q18				.467	
Q19					
Q20		.659			

Table B.3: Four factor pattern matrix

	Factor			
	1	2	3	4
Q1	.609			
Q2	.751			
Q3	.876			
Q4	.482			
Q5				
Q6	.672			
Q7				
Q8				
Q9			.979	
Q10			.426	
Q11				.426
Q12				.773
Q13				
Q14				.615
Q15			.668	
Q16		.786		
Q17		.992		
Q18		.468		
Q19				
Q20		.667		

Table B.4: Three factor pattern matrix

	Factor		
	1	2	3
Q1	.590		
Q2	.790		
Q3	.845		
Q4	.466		
Q5			
Q6	.656		
Q7	.430		
Q8			
Q9			.955
Q10			.498
Q11	.476		
Q12			
Q13			
Q14			.474
Q15			.716
Q16		.760	
Q17		.978	
Q18		.469	
Q19			
Q20		.664	

Appendix C

The following figures and tables refer to the initial CFA and structural models for the ordered model.

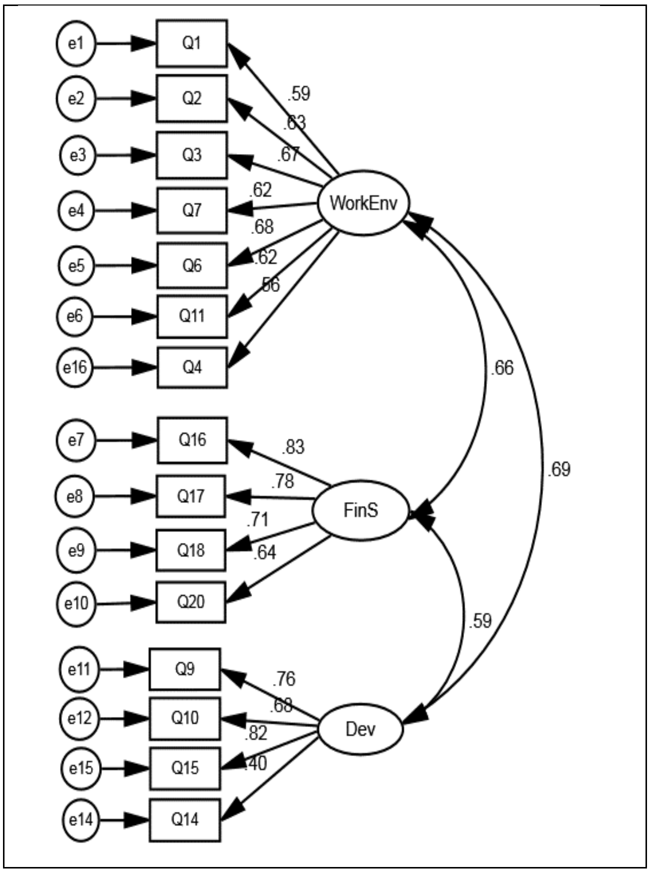


Figure C.1:Initial CFA model for ordered data

Table C.1: Regression weights for initial CFA ordered model

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
B1A11	<---	WorkEnv3	1			
B1A6	<---	WorkEnv3	1.145	0.113	10.171	***
B1A7	<---	WorkEnv3	1.091	0.114	9.588	***
B1A3	<---	WorkEnv3	1.238	0.122	10.126	***
B1A2	<---	WorkEnv3	1.133	0.118	9.63	***
B1A1	<---	WorkEnv3	1.051	0.115	9.128	***
B1A20	<---	FinS3	1			
B1A18	<---	FinS3	1.189	0.107	11.105	***
B1A17	<---	FinS3	1.458	0.123	11.831	***
B1A16	<---	FinS3	1.264	0.103	12.306	***
B1A4	<---	WorkEnv3	0.884	0.101	8.764	***
B1A15	<---	Dev3	1			
B1A10	<---	Dev3	0.752	0.06	12.594	***
B1A9	<---	Dev3	0.952	0.068	13.938	***
B1A14	<---	Dev3	0.513	0.072	7.129	***

Table C.2: Goodness-of-fit index values for initial CFA ordered model

Index	Index value
CFI	0.897
TLI	0.875
RMSEA	0.082
SRMR	0.067

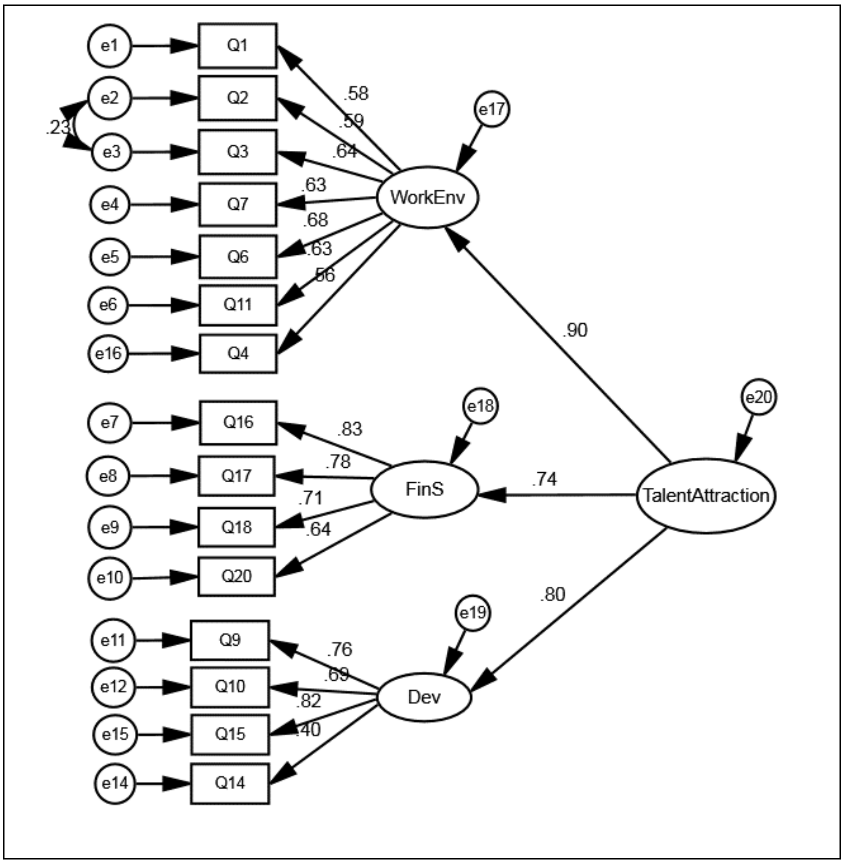


Figure C.2: Initial structural model for ordered model

Table C.3: Regression weights for initial structural ordered model

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
WorkEnv3	<---	TalentAttraction	1			
FinS3	<---	TalentAttraction	0.956	0.128	7.473	***
Dev3	<---	TalentAttraction	1.308	0.158	8.271	***
B1A11	<---	WorkEnv3	1			
B1A6	<---	WorkEnv3	1.139	0.111	10.258	***
B1A7	<---	WorkEnv3	1.087	0.112	9.68	***
B1A3	<---	WorkEnv3	1.162	0.119	9.737	***
B1A2	<---	WorkEnv3	1.053	0.115	9.141	***
B1A1	<---	WorkEnv3	1.035	0.113	9.128	***
B1A20	<---	FinS3	1			
B1A18	<---	FinS3	1.184	0.107	11.108	***
B1A17	<---	FinS3	1.457	0.123	11.857	***

B1A16	<---	FinS3	1.261	0.102	12.326	***
B1A4	<---	WorkEnv3	0.877	0.099	8.817	***
B1A15	<---	Dev3	1			
B1A10	<---	Dev3	0.758	0.06	12.654	***
B1A9	<---	Dev3	0.956	0.069	13.948	***
B1A14	<---	Dev3	0.516	0.072	7.145	***

Table C.4 : Goodness-of-fit index values for initial structural ordered model

Index	Index value
CFI	0.903
TLI	0.882
RMSEA	0.080
SRMR	0.065

Appendix D

The following figures and tables relate to the initial CFA and structural models for the randomised models.

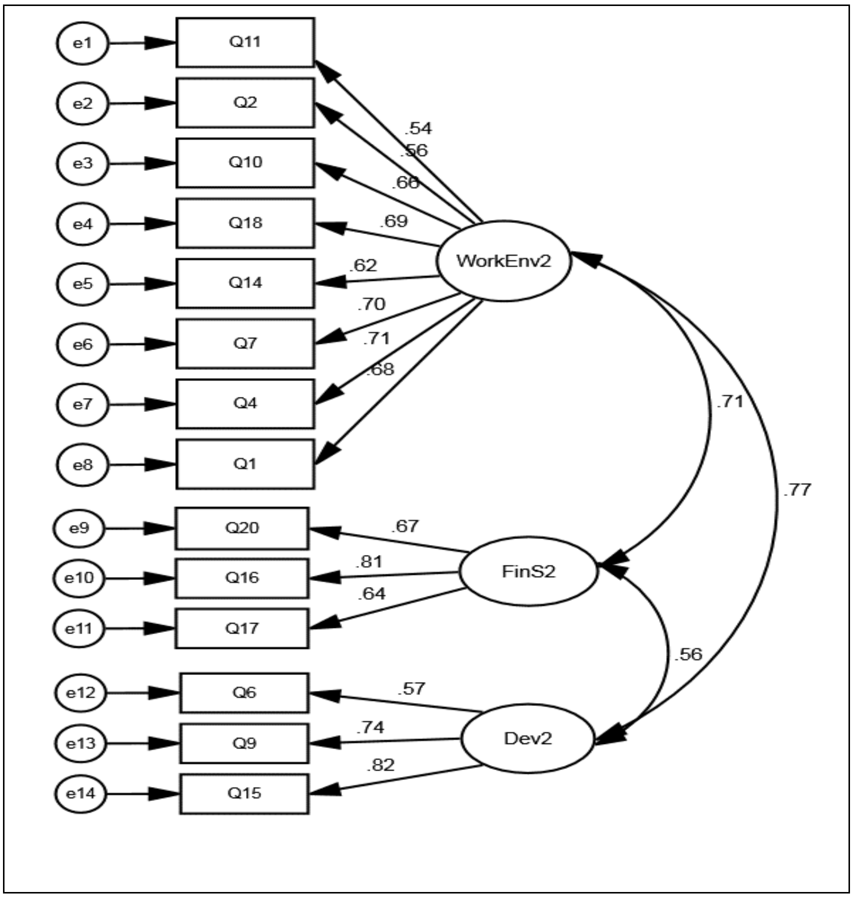


Figure D.1: Initial CFA randomised model

Table D.1:Regression weights for initial CFA randomised model

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
Q1	<---	WorkEnv2	1			
Q4	<---	WorkEnv2	0.963	0.055	17.414	***
Q7	<---	WorkEnv2	0.964	0.056	17.327	***
Q14	<---	WorkEnv2	0.952	0.062	15.403	***
Q18	<---	WorkEnv2	1.05	0.062	16.987	***
Q10	<---	WorkEnv2	0.874	0.054	16.327	***
Q2	<---	WorkEnv2	0.811	0.057	14.228	***
Q17	<---	FinS2	1			
Q15	<---	Dev2	1			
Q9	<---	Dev2	1.021	0.054	19.03	***
Q6	<---	Dev2	0.587	0.04	14.807	***

Q11	<---	WorkEnv2	0.787	0.058	13.682	***
Q16	<---	FinS2	1.017	0.064	15.966	***
Q20	<---	FinS2	0.867	0.059	14.674	***

Table D.2:Goodness-of-fit index values for initial CFA randomised model

Index	Index value
CFI	0.891
TLI	0.866
RMSEA	0.089
SRMR	0.052

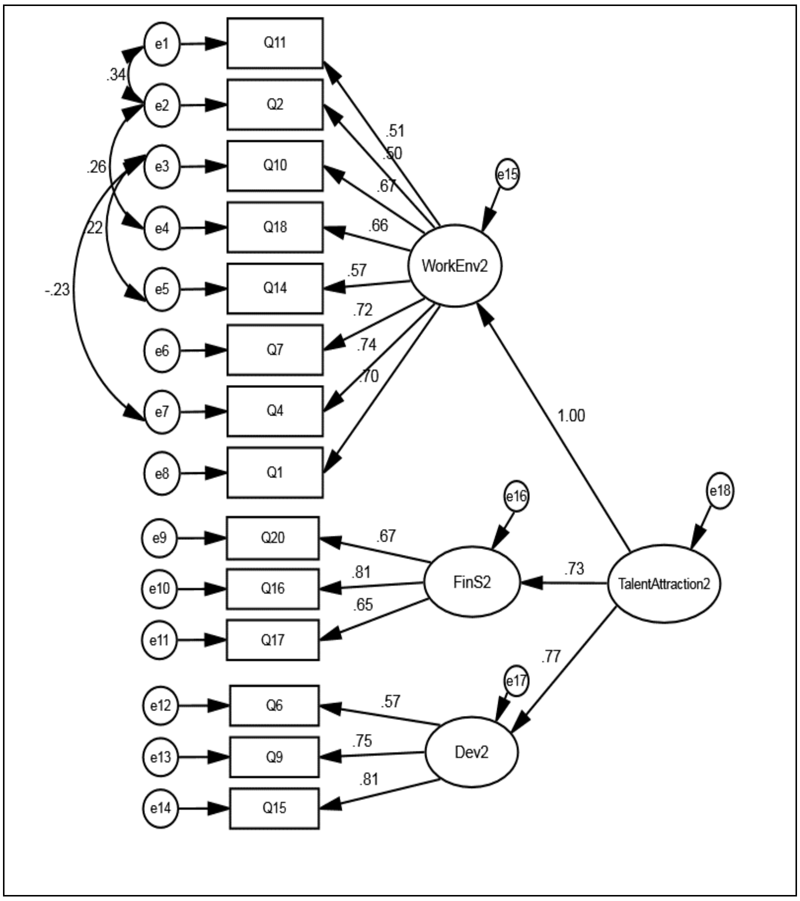


Figure D.2:Initial structural randomised model

Table D.3:Regression weights for initial structural randomised model

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
WorkEnv2	<---	TalentAttraction2	1			
Dev2	<---	TalentAttraction2	0.935	0.073	12.854	***
FinS2	<---	TalentAttraction2	0.699	0.062	11.207	***
Q1	<---	WorkEnv2	1			
Q4	<---	WorkEnv2	0.981	0.053	18.477	***
Q7	<---	WorkEnv2	0.952	0.053	18.09	***

Q14	<---	WorkEnv2	0.861	0.059	14.567	***
Q18	<---	WorkEnv2	0.983	0.058	16.859	***
Q10	<---	WorkEnv2	0.858	0.052	16.576	***
Q2	<---	WorkEnv2	0.693	0.055	12.702	***
Q17	<---	FinS2	1			
Q15	<---	Dev2	1			
Q9	<---	Dev2	1.027	0.054	19.023	***
Q6	<---	Dev2	0.587	0.04	14.748	***
Q11	<---	WorkEnv2	0.723	0.055	13.177	***
Q16	<---	FinS2	1.015	0.063	16.092	***
Q20	<---	FinS2	0.86	0.058	14.701	***

Table D.4:Goodness-of-fit index values for initial structural randomised model

Index	Index value
CFI	0.936
TLI	0.916
RMSEA	0.071
SRMR	0.044

Appendix E

The following table provides the regression weights for the ordered model imposed on the randomised data.

Table E.1:Regression weights for the ordered model imposed on randomised data

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
WorkEnv3	<---	TA3	1			
Dev3	<---	TA3	1.289	0.1	12.944	***
FinS3	<---	TA3	0.817	0.071	11.442	***
Q11	<---	WorkEnv3	1			
Q6	<---	WorkEnv3	0.888	0.073	12.212	***
Q7	<---	WorkEnv3	1.149	0.083	13.877	***
Q3	<---	WorkEnv3	1.115	0.084	13.27	***
Q2	<---	WorkEnv3	0.954	0.081	11.836	***
Q1	<---	WorkEnv3	1.243	0.089	13.919	***
Q20	<---	FinS3	1			
Q18	<---	FinS3	1.259	0.087	14.536	***
Q17	<---	FinS3	1.193	0.086	13.929	***
Q16	<---	FinS3	1.129	0.072	15.579	***
Q4	<---	WorkEnv3	1.223	0.085	14.455	***
Q10	<---	Dev3	0.67	0.043	15.744	***
Q9	<---	Dev3	1.005	0.055	18.345	***
Q15	<---	Dev3	1			

Appendix F

The following table provides the regression weights for the randomised model imposed on the ordered data.

Table F.1:Regression weights for randomised model on ordered data

Regression Weights: (Group number 1 - Default model)						
			Estimate	S.E.	C.R.	P
WorkEnv2	<---	TalentAttraction2	1			
Dev2	<---	TalentAttraction2	1.227	0.169	7.263	***
FinS2	<---	TalentAttraction2	1.244	0.18	6.908	***
Q1	<---	WorkEnv2	1			
Q4	<---	WorkEnv2	0.885	0.128	6.893	***
Q7	<---	WorkEnv2	1.256	0.156	8.056	***
Q14	<---	WorkEnv2	0.717	0.139	5.142	***
Q18	<---	WorkEnv2	1.58	0.185	8.533	***
Q10	<---	WorkEnv2	1.302	0.154	8.429	***
Q17	<---	FinS2	1			
Q15	<---	Dev2	1			
Q9	<---	Dev2	0.964	0.072	13.305	***
Q6	<---	Dev2	0.67	0.065	10.308	***
Q11	<---	WorkEnv2	1.13	0.142	7.931	***
Q16	<---	FinS2	0.888	0.061	14.67	***
Q20	<---	FinS2	0.707	0.058	12.109	***